INFRARED SPECTROSCOPY FOR QUALITATIVE AND QUANTITATIVE ANALYSIS OF MANGOES (*Mangifera indica*): A CHEMOMETRIC APPROACH

Thesis Submitted for the Award of the Degree of

DOCTOR OF PHILOSOPHY

in

Physics

By

Agnibha Das Majumdar

Registration Number: 11915455

Supervised By

Dr. Neha Munjal (18869)

Department of Physics (Associate Professor)

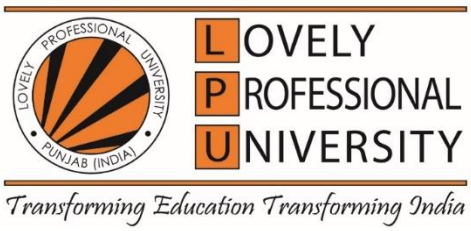
Lovely Professional University, Punjab

Co-Supervised by

Dr. Uma Kamboj (21636)

Department of Physics (Assistant Professor)

Lovely Professional University, Punjab



LOVELY PROFESSIONAL UNIVERSITY, PUNJAB

2023

DECLARATION

I, hereby declared that the presented work in the thesis entitled "**Infrared Spectroscopy For Qualitative and Quantitative Analysis of Mangoes (*Mangifera Indica*): A Chemometric Approach**" in fulfillment of degree of **Doctor of Philosophy (Ph. D.)** is outcome of research work carried out by me under the supervision of Dr. Neha Munjal, working as assistant professor, in the department of Physics, school of Chemical Engineering and Physical Sciences of Lovely Professional University, Punjab, India. In keeping with general practice of reporting scientific observations, due acknowledgements have been made whenever work described here has been based on findings of other investigator. This work has not been submitted in part or full to any other University or Institute for the award of any degree.

(Signature of Scholar)

Name of the scholar: Agnibha Das Majumdar

Registration No.: 11915455

Department/school: Department of Physics, School of Chemical Engineering and Physical Sciences, Lovely Professional University, Punjab, India

CERTIFICATE

This is to certify that the work reported in the Ph. D. thesis entitled "**Infrared Spectroscopy For Qualitative and Quantitative Analysis of Mangoes (*Mangifera Indica*): A Chemometric Approach**" submitted in fulfillment of the requirement for the reward of degree of **Doctor of Philosophy (Ph.D.)** in the department of Physics, School of Chemical Engineering and Physical Sciences, is a research work carried out by Agnibha Das Majumdar, 11915455, is bonafide record of his original work carried out under my supervision and that no part of thesis has been submitted for any other degree, diploma or equivalent course.

(Signature of Supervisor)
Name of supervisor: Dr. Neha Munjal
Designation: Assistant Professor
Department of Physics
Lovely Professional University
Phagwara, Punjab

(Signature of Co-Supervisor)
Name of Co-Supervisor: Dr. Uma Kamboj
Designation: Assistant Professor
Department of Physics
Lovely Professional University
Phagwara, Punjab

My Ph.D. thesis is dedicated to my loving parents,

my soulmate and my respected guide

ABSTRACT

Mango (*Mangifera indica*) is one of the most commonly consumed tropical fruits with enormous nutrients. It is having a commercial demand throughout the world. India has number of varieties, with which different food and beverage industries have been grown up. It is very important to know the nutritional parameters of mango in terms of human health and nutrition. “King of fruits”, mango is one of the popular and economically accepted tropical fruits too. It has different nutrients (mainly macronutrients and micronutrients) and this fruit can be taken in form of intact, as well as in the form of many by-products. Many of the food products like juice and soft drinks (beverages), puree, nectar, mango leather is formed from the mango pulp. For the market value it is essential to know the pulp quality of the mango fruit. Mango pulp is a viscous fluid and to know the transport property the viscosity is the major parameter. Food science and food processing systems are having a big role of the transport property of the fluid. Thus, the present study aims to investigate the different physical, nutritional and rheological parameters. For this, fifty one intact ripe mature mango samples have been collected from the different regions of India namely Bihar, Maharashtra, Uttar Pradesh, Punjab and West Bengal. For physical parameter analysis initially three axes (major, minor and intermediate) has been determined and with some mathematical equations physical parameters like Sphericity, surface area and size has been calculated. Peel, pulp and kernel weight has been examined and compared the data to know the maximum presence of quantity in an intact fruit. Estimation of chemical properties was done with standard methods of Association of Official Agricultural Chemists (AOAC). Different parameters of Indian mango pulp like carbohydrate, crude protein, crude fiber, crude fat, moisture, ash content, Total Soluble Solid (TSS) and energy value (in Kcal) has been determined and The viscosity is calculated using cup method and analysis of the same has been done through descriptive statistics.

The correlation between all those physical and chemical parameters has been analyzed.

Energy is negatively and positively correlated with moisture and carbohydrate with R^2 value of 0.919 and 0.936 respectively and carbohydrate is negatively correlated with moisture with R^2 value of 0.861. Intact fruit weight is positively correlated with pulp content, size of intact mango and surface area of intact mango with R^2 value of 0.977, 0.945 and 0.962 respectively. The pulp content is also positively correlated with the size of intact mango and surface area of the intact mango with R^2 value of 0.912 and 0.931 respectively. There is also a positive correlation between size and surface area of intact mango with R^2 value of 0.997. Also a statistical relation between energy, moisture, carbohydrate and pulp content, weight, size and surface area of intact fruit has been developed for the mango pulp.

Near infrared spectra and Fourier Transform infrared spectra has been collected to identify the infrared active molecules present in the mango pulp with the NIR FOSS2500 NIR spectrometer between the spectral range of 700-2500 nm and Perkin Elmer FTIR spectrometer with KBr pallets between the spectral ranges of 4000-700 cm^{-1} respectively. Principal Component Analysis (PCA) has been performed on all physical, chemical and spectral (near infrared and Fourier Transform infrared) data to know the trend of the data. Singular value decomposition with imputation has been used to determine the principle components (PCSs) within the Clustvis web tool application. All the results have been compared. The sample collected from Uttar Pradesh and West Bengal has been classified well in form for all the properties.

The present study also has established a green, robust, non-destructive, time effective and cost effective method to predict the quality of mango using near infrared spectroscopy together with chemometric approach. The spectra has been recorded for the fifty one different homogenized Indian mango pulp within the range of 700-2500 nm and then wavelength selection method has been applied to reduce the wavelengths for prediction. Multiplicative Scatter Correction (MSC) and second order derivative (with second order polynomial and three smoothing points) were used to pre-process of the original spectra and Partial Least Square regression method has been used to develop the prediction model for sweetness analysis of mango. During the partial Least Square regression, for

getting more accurate results wavelength selection methods (i.e. interval Partial Least Square regression and synergic Partial Least Square regression) have performed. Coefficient of determination for calibration (R_c^2) as well as for validation (R_v^2) has been given the results more than 0.9 for every qualitative parameter. For every case, a result of maximum factor 7 has been taken and it has shown that at factor the values of R_c^2, R_v^2, RMSEC and RMSEV are best. For the Partial Least Square regression the value of coefficient of determination for calibration and the value of root mean square error for calibration (RMSEC) for moisture is presented as 0.9937 and 0.0033 and coefficient of determination for validation and root mean square error for validation (RMSEV) has been presented as 0.9907 and 0.0040 respectively. The value of coefficient of determination for calibration and the value of RMSEC for crude fat is presented as 0.9619 and 0.0458 and coefficient of determination for validation and RMSEV has been presented as 0.9492 and 0.0530 respectively. The value of coefficient of determination for calibration and the value of RMSEC for crude protein is presented as 0.9822 and 0.0146 and coefficient of determination for validation and RMSEV has been presented as 0.9744 and 0.0175 respectively. The value of coefficient of determination for calibration and the value of RMSEC for crude fiber is presented as 0.9712 and 0.0163 and coefficient of determination for validation and RMSEV has been presented as 0.9590 and 0.0194 respectively. The value of coefficient of determination for calibration and the value of RMSEC for pH is presented as 0.9913 and 0.0085 and coefficient of determination for validation and RMSEV has been presented as 0.9868 and 0.0105 respectively. The value of coefficient of determination for calibration and the value of RMSEC for carbohydrate is presented as 0.9923 and 0.0197 and coefficient of determination for validation and RMSEV has been presented as 0.9892 and 0.0234 respectively. The value of coefficient of determination for calibration and the value of RMSEC for total soluble solids is presented as 0.9576 and 0.0270 and coefficient of determination for validation and RMSEV has been presented as 0.9426 and 0.0315 respectively. For iPLS and siPLS the results has modified in terms of coefficient of determination and root mean square errors. Finally the siPLS has given the best results with minimum wavelengths. siPLS has

reduced the wavelength range of determination from 700-2500. For estimation of moisture the important wavelength ranges are 1300-1500 and 2000-2500, for crude fat the important wavelength ranges are 1100-1300 and 1500-2300, for crude protein the important wavelengths are 1900-2100 and 2300-2500, for crude fiber the important wavelength ranges are 1900-2100 and 2300-2500, for pH content the important wavelength ranges are 1500-1700 and 2100-2500, for carbohydrate the important wavelength ranges are 1500-1700 and 2300-2500 and for total soluble solid the important ranges are 1300-1500 and 1700-1900. For the synergic interval Partial Least Square regression the value of coefficient of determination for calibration and the value of RMSEC for moisture is presented as 0.9569 and 0.0093 and coefficient of determination for validation and RMSEV has been presented as 0.9529 and 0.0153 respectively. The value of coefficient of determination for calibration and the value of RMSEC for crude fat is presented as 0.9961 and 0.0032 and coefficient of determination for validation and RMSEV has been presented as 0.9044 and 0.0021 respectively. The value of coefficient of determination for calibration and the value of RMSEC for crude protein is presented as 0.9901 and 0.0032 and coefficient of determination for validation and RMSEV has been presented as 0.9155 and 0.0197 respectively. The value of coefficient of determination for calibration and the value of RMSEC for crude fiber is presented as 0.9752 and 0.0065 and coefficient of determination for validation and RMSEV has been presented as 0.9107 and 0.0149 respectively. The value of coefficient of determination for calibration and the value of RMSEC for pH is presented as 0.9800 and 0.0164 and coefficient of determination for validation and RMSEV has been presented as 0.9553 and 0.0196 respectively. The value of coefficient of determination for calibration and the value of RMSEC for carbohydrate is presented as 0.9345 and 0.0338 and coefficient of determination for validation and RMSEV has been presented as 0.9145 and 0.0389 respectively. The value of coefficient of determination for calibration and the value of RMSEC for total soluble solids is presented as 0.9513 and 0.0245 and coefficient of determination for validation and RMSEV has been presented as 0.9344 and 0.0249 respectively. From the results it is indicated that NIR spectroscopy along with

chemometrics is giving the promising results and is an efficient method to calculate the quality parameters of mango fruit with green and non-destructive way.

ACKNOWLEDGEMENT

I would like thank with all the respect to the Almighty who has showered blessings on me to get through the complete task of my Ph.D. research work. I wish to thank all those who made this thesis possible.

It gives me an immense debt of gratitude to express my sincere thanks to my supervisor, **Dr. Neha Munjal**, Assistant Professor, Department of Physics, Lovely Professional University, Phagwara, for providing me the best opportunity to carry out my research under her excellent guidance. I am very grateful for her continuous encouragement, whole-hearted co-operation, motherly care and patience throughout the course of Ph.D. I feel very fortunate to get a chance to work under her patient supervision. I have learnt the time management and work management from her. Her continuous motivation towards research made me a good researcher and helped me a lot be a presentable researcher.

My special thanks to my co-supervisor, **Dr. Uma Kamboj**, Assistant Professor, Department of Physics, Lovely Professional University, Phagwara, for providing me the best guidance to carry out my research with patients. Her critical and valuable comments always motivate me for betterment.

I would like to convey my special gratitude to my Head of the School, School of Chemical Engineering and Physical Sciences, Lovely Professional University, Punjab, **Prof. Kailash Chandra Juglan**, Professor, Department of Physics, for providing me the opportunity to conduct my research within his school and also for giving me always the positive response and motivation during my Ph.D. research work.

Also, I would like to convey my special thanks to **Dr. Shiv Singh**, Assistant Librarian, Bennett University, Greater Noida, Uttar Pradesh, for providing me all Scopus and Web of Science data for my literature review. His continuous motivation towards solving the research problems made me a good thinker and helped me a lot as a good data analyst.

I feel very fortunate to get a chance to work under their patient supervision. I am

highly obliged to **Dr. Gautam Pramanik**, Scientist-D, UGC DAE Consortium Scientific Research Centre, Kolkata, and **Dr. Sunita Mishra**, Senior Principal Scientist, CSIR-CSIO, Chandigarh, for not only allowing me to carry out the experimental work in their labs but also for their exceptional guidelines, critical comments and providing an excellent atmosphere for research work. I specifically thankful to Dr. Pramanik for his deep insights and expertise helped me throughout the various stages of my research.

I am extremely grateful to **Dr. Sutapa Bhowmick**, Head, Advanced Analytical Testing Laboratory, for allowing me to carry out my analytical experimental analysis in her lab and for her obligated support, continuous incitement and valuable suggestions during my Ph.D.

I feel very blessed to have **Ms. Pavas**, my senior research scholar, Lovely Professional University and my dear colleagues **Ms. Bhumandeep Kaur, Mrs. Amandeep Kaur, Ms. Enakshi Dutta, Mr. Diptanu Banik, Mr. Harminder Singh**. I remain indebted for their kind sustenance during my personal difficult times during my Ph.D. I am also thankful to **Mr. Pradipta Samanta**, Senior Project Fellow, CSIO-Chandigarh, for supporting me in the experimental work during my Ph.D.

Finally, I would like to acknowledge the most important persons of my life – my parents, **Mrs. Tanusree Das Majumdar** and **Mr. Anupam Das Majumdar** who has been the pillar to my strength, the encouragement to my goals, and the root to which I have based my life on. I am indebted to them for always showing me the light at the end of the tunnel. They have taught me to value myself, filled me with courage and confidence at every stage of life.

A special thanks to my beloved best friend and would be life partner **Ms. Poulamee Chackrobarty**, for her love, affection and moral support throughout my Ph.D. work.

Agnibha Das Majumdar

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AOAC	Association of Agricultural Chemists
ANOVA	Analysis of Variance
ASCA	ANOVA Simultaneous Component Analysis
BLC	Baseline Linear Correction
BP-ANN	Back Propagation- Artificial Neural Network
BSV	Black Streaked Vascular
CSIR-CSIO	Central Scientific Instrumental Research-Central Scientific Instrumental Organization
CARS	Competitive Adaptive Reweighted Sampling
DOP	Dynamic Orthogonalization Projection
DM	Dry Matter
DA	Discriminant Analysis
DL	Deep Learning
DAF	Day After Flowering
DTGS	Deuterated Triglycine Sulphate
FT-NIR	Fourier Transform Near Infrared
FTIR	Fourier Transform infrared
GC-MS	Gas Chromatography- Mass Spectroscopy
GA	Genetic Algorithm
HPLC	High Performance Liquid Chromatography
IBD	Internal Breakdown
iPLS	Interval Partial Least Square regression

IBM SPSS	International Business Machine Statistical Package for Social Sciences
IR	Infrared
IFR	Insidious Fruit Rot
KP model	Kensington Pride model
KNN	K-Nearest Neighbor
LSSVM	Least Square Support Vector Machine
LDA	Linear Discriminant Analysis
LW	Loading Weighs
MCT	Mercury Calcium Telluride
MLR	Multiple Linear Regression
MSC	Multiplicative Scatter Correction
NIR	Near infrared
NIRS	Near infrared spectroscopy
NIPLAS	Non-linear Iterative Partial Least Square
NIRED	Near Infrared Emitting Diode
OAV	Odour Activity Value
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS	Partial Least Square regression
PCA	Principal Component Analysis
PC	Principal Component
QDA	Quadratic Discriminant Analysis
R_c^2	Coefficients of determination for calibration
R_v^2	Coefficients of determination for validation
RMSEC	Root mean square error for calibration
RMSEV	Root mean square error for validation
RC	Regression Coefficient
SVD	Singular Value Decomposition

SPME	Solid Phase Micro Extraction
SSC	Soluble Solid Content
SG	Savitzky-Golay
SPA	Successive Projection Algorithm
SVM	Support Vector Machine
SRP	Sample Regression Plane
SVD	Singular Value Decomposition
siPLS	Synergy Partial Least Square regression
TSS	Total Soluble Solid
TA	Titratable Acidity
UV	Ultra Violet
UV-Vis	Ultra Violet - Visible
UGC-DAE	University Grant Commission- Department of Atomic Energy

OUTLINE OF THE THESIS

Present Ph.D. Research work focused on the estimation of quality parameters of mango using NIR spectroscopy and chemometrics.

Chapter-1 includes the introduction of the basic topic and techniques which are used in the present research work. Starting from electromagnetic spectrum to the different types of spectroscopy, the thesis then covers about the basic concept of NIR spectroscopy with its history and basic principal and multivariate analysis. An introduction about the mango is given in brief.

Chapter-2 covers the literature review and highlighted research works which is already carried to determine the different destructive and non-destructive studies of mangoes.

Chapter-3 covers research gap obtained from the literature survey and following objectives of the present research work. This chapter also focuses on the research methodology used during the research. Starting from sample collection, sample preparation, physical, chemical and rheological parameter characterization, statistical methods employed in this research have been listed in this chapter.

Chapter-4 covers results obtained after the characterization of physical, chemical and rheological parameters of mango using standard methods.

Chapter-5 covers the results obtained for the classification of above mentioned parameters after applying the Principal Component Analysis.

Chapter-6 covers the development of statistical regression model to estimate the quality parameters of mangoes.

Chapter-7 and Chapter-8 has followed the conclusion and future scope of the present study respectively.

Chapter-1: Introduction

Chapter-1

Introduction

1.1.Spectroscopy:

The field of physics known as spectroscopy is concerned with the observation and analysis of radiation emitted or absorbed by atoms and molecules in order to provide information on their structural composition. It provides information not only about the arrangement and motion of the outer electron (optical spectroscopy), but also about the inner electrons (X-Ray spectroscopy) and their angular momentum, magnetic moment, distribution of charge and magnetism of the nucleus. In 1672, it was Sir Isaac Newton who originated the term "spectroscopy" by showing that prism refracts blue light more than red light, thus forming a band of colors known as 'spectrum'. It is known that the visible spectrum is a very small part of the broad electromagnetic spectrum (Figure. 1.1) which ranges from the γ-ray to the radio waves. The wavelength λ and the frequency ν of a monochromatic wave is related as,

$$\lambda \nu = v \tag{1}$$

Where, v is the phase velocity in the medium. The phase velocity v of an electromagnetic wave in a medium is related to the phase velocity c in a vacuum by,

$$v = c / \mu \tag{2}$$

Where, μ is the refractive index of the medium with respect to vacuum [1]

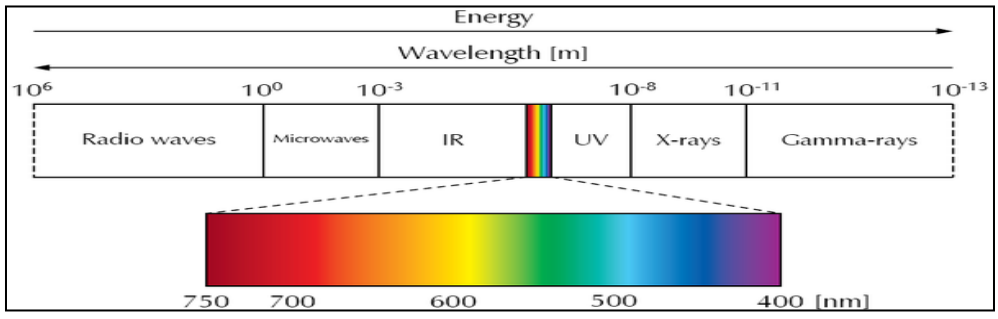


Figure 1. 1 Electromagnetic Spectrum
Source: [2]

The term "electromagnetic spectrum" refers to the arrangement of all forms of electromagnetic radiation in descending frequency or increasing wavelength order. The prime focus is on following parts of the electromagnetic spectra:

(a) The wavelength range of **visible and ultraviolet radiations** spans 0.01 to 0.8 μm. In a conjugated or unconjugated system, the excitation of an electron results from the absorption of radiation in this area. Because there is less distance between two different states that is ground and excited in a conjugated system, absorption happens at a longer wavelength. Additionally, the carbonyl group of an aldehyde or ketone absorbs light at specific wavelengths. Thus, a visible or ultraviolet spectrum may not reveal any information about the remaining portion of the molecule but is highly useful for detecting conjugation, carbonyl groups, etc.

(b) The Infrared radiations lie over the range of wavelength from 700 nm to 1 mm (near, mid and far infrared region). These radiations are having higher values in terms of wavelengths and therefore showing lower energetic from the UV-Visible radiations. The molecular vibration occurs with the absorption of the infrared radiations by an organic compound. The transitions between the vibrational levels are followed by the transitions between the rotational levels. Thus, concern bands appear which purposively absorb for the stretching vibrations and are very useful for elucidating the structure of any compound. The absorptions at higher wavelength in the infrared region are having most characteristics of a compound and also help in differentiating the compounds from each other. Detection of the functional groups (C=O, O-H, NH_2 etc.) with specific vibrational frequencies can be studied within this region. Although, NIR radiation are more useful than ultra-violet radiation but, it cannot explain the full information about the environmental effects in a molecule.

It is possible to ascertain the structure of an unknown compound from its ultraviolet and infrared spectra [3].

There are broadly two branches of spectroscopy: -

1.1.1 Atomic Spectroscopy:

In atoms, when atom absorbs the energy the electrons get excited and jump to the higher

excited state and due to instability of electrons in excited state these come back in ground state and emission of energy occurs which is equal to the difference between two energy levels. This helps in the determination of the structure of atoms.

This kind of spectroscopy has been used to determine the organic and inorganic molecules present in the metallic atoms and the study is based upon the principle of atomic emission and absorption spectroscopy. Two types of energy that is thermal and radiation energy are considered as the incident energy. On the implementation of the incident energy the electrons are excited from ground energy state to excited state. This phenomenon is known as absorption spectroscopy and when the electron jumped from excited energy state to ground energy state, it is known as emission spectroscopy. The radiation of UV-Vis spectrum is responsible for electronic transition. The electron from the valence shell of the atom get excited and then this spectroscopy is sensitive to trace the chemical and physical properties of any element.

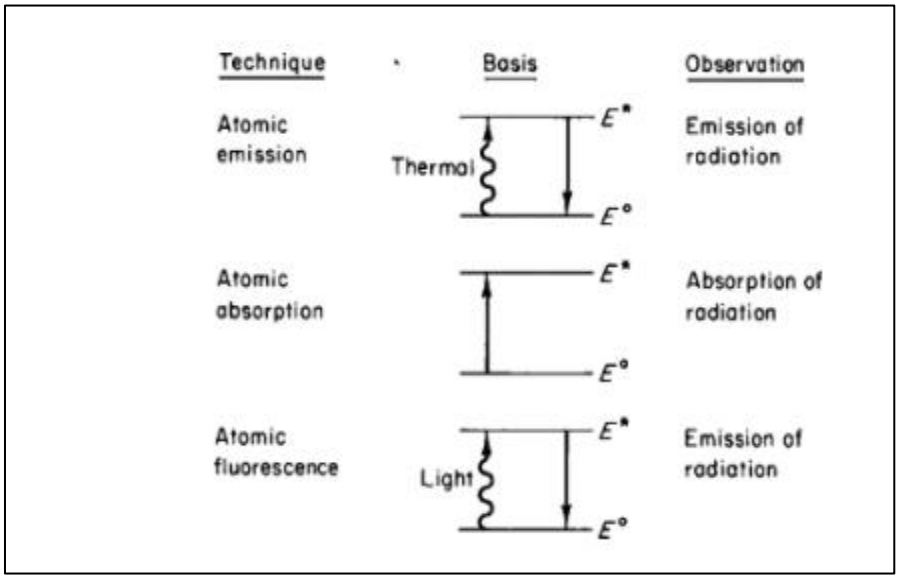


Figure 1. 2 Types of Atomic Spectroscopy
Source: [4]

1.1.2 Molecular Spectroscopy:

In the study of molecules, the energy is absorbed by molecules result, in rotational, vibrational and electronic transitions which are depending upon the amount of energy absorbed. Rotational, vibrational and electronic energy levels of molecule are known

as molecular energy levels. The transitions of energies taking place between these levels results in the molecular spectrum.

The area above the visible spectrum is known as the infrared (IR) region, and the area below is known as the ultraviolet (UV) spectrum radiations. Compared to the visible region, infrared light has a longer wavelength and is less energetic. When light is going through an organic compound sample, certain of the light's wavelengths are absorbed while others from the same light source are unaffected. Atomic electrons are energized when light radiation travels through an organic substance. A molecule absorbs the light according to its molecular vibrations, overtones and harmonics [5].

The goal of molecular spectroscopy is to identify the frequencies of radiation that a specific substance absorbs or emits, and to attempt to trace the patterns of energy absorption or emission to specifics of molecule structure. According to quantum mechanics, electromagnetic radiations are thought to travel in distinct energy packets known as photons. These photons are quantized and have highly definite energy. A photon's energy is determined by the relationship,

$$E = h \nu \tag{3}$$

Where, E is the energy, h is plank’s constant and ν is the velocity of the medium. The energy states of a molecule are quantized. By exposing an atom or molecule to electromagnetic radiation with a frequency equal to the energy difference between states E1 and E2, one can induce an atom or molecule to move from a lower energy state E1 to a higher energy state E2. Equal energy is released when the atoms or molecules transition back from condition E2 to state E1. Thus, certain electromagnetic radiation quanta must be absorbed in order to excite an organic molecule. A molecule can only absorb a photon of electromagnetic radiation if its energy exactly matches the difference between two of its states. A specific frequency or wavelength of electromagnetic radiation will be produced by the energy differential between excited and ground state, depending on the type of the transition and the frequency is given by [6],

$$\Delta E = h \nu = hc/\lambda \tag{4}$$

Where, ΔE is the change id energy, h is the plank's constant, v is the velocity of the medium, c is the velocity of light and λ is the wavelength.

1.1.3 Types of Molecular Spectra:

When light from a source containing a substance in "molecular" state is sent into a spectrometer, more or less broad wavelength regions are observed in the spectrum. These regions are called 'bands' and the spectrum is called 'band spectrum' or 'molecular spectrum'.

There are three spectral ranges of the molecular spectra correspond to the various kinds of transitions between the molecular energy levels.

(a)Electronic Spectra: In the ultra-violet and visible region both the electronic transitions that is for emission and absorption has been observed. Each spectrum has included number of bands with sharp edges. Band-head is the name given to these pointed edges. During the band's head, the intensity abruptly drops to zero, and from this edge, it keeps increasing on the opposite side of the band. These band heads are discovered to be a collection of spectral lines with very high resolution that grow farther apart the more away they are. According to legend, the band deteriorates toward the side that is opposite the band head. For both the heteronuclear and homonuclear molecules the electronic transitions taking place and therefore electronic spectra has observed.

(b)Vibrational-Rotational Spectra: Vibrational spectra are described mainly for the absorption of the NIR radiation by the heteronuclear molecules. Each spectrum is made up of intense bands known as 'fundamental' bands which tend to be very weak bands. With very fine structure, homonuclear molecules (such as H_2, N_2, and O_2) do not form vibrational-rotational bands; instead, the vibrational-rotational spectra are only detected for heteronuclear molecules.

(c)Pure Rotational Spectra: During the absorption of the radiation of far and microwave region pure rotational spectra is observed. Each spectrum is made up of a series of spectral lines and the distance between those lines are equal. Only the heteronuclear diatomic compounds like HBr, HF, CO, HCl, etc. exhibit pure rotational

spectra. The simple succession of absorption maxima that are closely equidistant on a wavenumber scale make up the rotating spectra. For homonuclear molecules, which don't give rotational spectra, the moment of inertia and the inter-nuclear distance can be found from the rotational fine structure of electronic bands.

The origin of different molecular spectra lies in the different energy states of the molecule.

1.1.4 The Molecule as a Harmonic Oscillator:

The nuclei in the diatomic molecule in the diatomic molecule vibrate along the inter-nuclear axis, and this is responsible for the near infrared spectra. Treating the molecule as a harmonic oscillator is the most basic assumption that can be made regarding the nature of vibrations. The potential energy function that causes the nuclei to vibrate is parabolic in this case and is given by

$$V(r) = 1/2k(r-r_e)^2 = 1/2kx^2 \quad (5)$$

Where, k is the force constant, r_e is the equilibrium inter-nuclear distance, and x is the displacement of the oscillator from the equilibrium position. After solving the Schrödinger equation it can be derived that

$$E = h\nu\,(v+1/2) \quad (6)$$

Where, E is energy, h is plank's constant, ν is the frequency. This gives the allowed energies for the harmonic oscillator and v is the vibrational quantum number which can take the integral values like 0, 1, 2, 3 ……

1.1.5 The Molecule as Anharmonic Oscillator:

An major discrepancy can be seen when comparing an observed near-infrared spectrum to that predicted by a diatomic molecule regarded as a harmonic oscillator. The harmonic oscillator would give a single band at wave number ω, which is the classical frequency of vibration of the molecule. The actual infrared spectrum is, however found to consists of an intense (fundamental) band at ω, plus a number of weak bands (overtones) at wave numbers slightly lesser and lesser than 2ω, 3ω and so on. The observations of the overtones indicates that the selection rule $\Delta v = \pm 1$ is not strictly obeyed, and transitions corresponding to $\Delta v > 1$ do take place. This in turn, is

attributed to the fact that the dipole moment of the molecule is not strictly linear with respect to the inter-nuclear displacement x. This is expressed as "electrical anharmonicity" of the molecule. The observation that the overtones appear not exactly at 2ω, 3ω … but at lesser and lesser values focuses about the vibrational energy levels are not exactly equally spaced, but converge slowly. This is attributed to the fact that for an actual molecule the potential energy function is not harmonic in nature. This is expressed as "mechanical anharmonicity" of the molecule. The energy values of the anharmonic oscillator are given by,

$$E = hc\omega_e (v+1/2) - hc\omega_e x_e (v+1/2)+\ldots.. \quad (7)$$

The quantity ω_e is the wave number spacing of energy levels that would occur if the potential curve were a parabola, $\omega_e x_e$ is the 'anharmonicity constant' which is much smaller than ω_e and is always positive [1].

1.2.Infra-red Spectroscopy:

One of the most used strategies for identifying functional groups in pure compounds, mixtures, and when contrasting different compounds is this one. Infrared spectroscopy is the technique which can replace many conventional analytical methods that is Gas chromatography, High pressure liquid chromatography Nuclear Magnetic Resonance spectroscopy and Mass spectroscopy. Those analytical methods need sample preparation and involve many chemicals for qualitative and quantitative analysis of cosmetic products, food and beverages, many agricultural products, polymers and petrochemicals etc. But, infrared spectroscopy has introduced a cost and time effective measurement tool with ecofriendly and specific approach. Atoms are capable of rotating around solitary covalent bonds. Moreover, covalent bonds can vibrate with stretching and bending just like the concern atoms were joined by flexible springs. Infrared spectroscopy leads different vibrations of organic molecules. The energies associated with transitions between vibrational levels in organic molecules correspond to frequencies in the infrared region which stretches from 400-4000 cm^{-1} (Mid Infrared region) and 4000-12500 cm^{-1} (Near Infrared region) [7].

Since it provides enough details on a compound's structure, the infrared spectrum is a

significant component of the electromagnetic spectrum. Unlike to the ultraviolet spectrum, which has just a few number of peaks, IR spectroscopy produces a spectrum with several bands of absorption from which qualitative information about the structure of any organic chemical can be extracted. The different bands in a molecule stretch and flex relative to one another as a result of infrared radiation absorption. The range from 0.75 µm to 2.5 µm, 2.5 µm to 25 µm, and 25 µm to 1000 µm is referred to as the Near Infrared, Mid Infrared, and Far Infrared, respectively [3].

1.2.1. Historical Development of Infra-red Spectroscopy

It is the time of second world war there was a huge science revolution was taking place and several physical research techniques has been introduced against the traditional chemical research. Infrared spectroscopy was one of them. Infrared spectroscopy was achieved its highest popularity between the time of 1950 to 1960. Although in the present decade it has become very popular instrumental method with application in a broad sector. Structural diagnosis (qualitative analysis) is the primary function of the infrared spectroscopy, although not only qualitative analysis but also quantitative analysis has taken the place for infrared spectroscopic technique. Refractometry, densitometry, boiling point analyzer were the different discrete methods used to know about the nature of molecules present in the compound, which is not quite precise and time friendly. As an alternative infrared spectroscopy is quite rapid and accurate technique. This is also an advantage of this technology that tested sample can be of any form that is liquid, solid, gas, powder or thin films. To accomplish this technique an expert of this field Norman Wright explained in 1966 that, “From a little known and relatively unused method two and half decades ago, infrared spectroscopy has now reached a status of general acceptance and utility unequaled among the analytical methods of organic chemistry.” It had become more popular when in 1948 the first text was published on this field named as “Practical Spectroscopy”.

In the year of 1800 Sir William Herchel has accidentally discovered the first non-visible spectra, which he has named as Infrared spectroscopy as it was beyond the red spectrum. He has explained the experimentation with the help of glass prism and blackened

thermometer. Thermometer has shown a deviation in temperature when it falls into the region beyond the red light of solar spectrum. Then, in the year of 1882 Abney and Festing proceeded further and invented the IR spectrum and obtained the spectra of 50 groups of molecules. After that Julius has invented the spectra for more 20 groups, specifically the methyl group. In the year of 1903 Sir W.W. Coblentz has invented over 100 organic groups and made a series of IR spectrum. Now, there is need to produce an instrument which can calibrate and construct the IR spectra with easiest way. Then in the time of Second World War United States Government has shown an interest in the development of that kind of instrumentation. Finally the company named Perkin-Elmer has invented and constructed the first model of infrared spectrometer. After that many modifications has been done till the date [8, 9].

1.2.2. Principle of Infrared Spectroscopy:

Food safety is presently a prime concern of human civilization. It needs to be maintained for good health and nutritional life. There are several analytical tools which have been used for last couple of decades to maintain the food safety and quality control. The methods such as GC-MS and HPLC and Nuclear Magnetic Resonance (NMR) Spectroscopy are quite time consuming and no specific methods. So, to overcome their drawbacks and dis-advantages different non-destructive spectroscopic methods like Near Infrared Spectroscopy (NIR), Fourier Transform Infrared Spectroscopy (FTIR), Raman Spectroscopy, and Fluorescence Spectroscopy etc. has been widely used in the field of food technology. Among those NIR, along with chemometrics, has been taken an important and versatile role to develop several statistical models and devices for quality assurance of different fields. A molecule absorbs the infrared with the matching frequency of vibration and excite from lower energy level to higher energy level (vibrational levels). There are number of rotational levels exists between every vibrational levels. As after absorbing the radiation the molecule vibrates and also rotates on its axis, the infrared spectra have shown both vibrational-rotational spectra. Only those bonds in a molecule that are followed by a change in dipole moment can absorb infrared energy; all other bonds in a molecule

cannot. These vibrational transitions, which also involve a modification of the molecule's dipole moment, are known as infrared active transitions. Therefore, these are in charge of absorbing radiation in the infrared zone. On the other hand, the vibrational transitions that do not result in a change in the molecule's dipole moment are not directly visible and are inactive in the infrared spectrum. For instance, the dipole moment varries along with vibrational transitions in the C=O, N–H, O–H, and other bands, which substantially absorb in the infrared range. However, symmetrical alkenes and alkynes do not absorb in the infrared region because the transition in the C-C bond is not followed by a change in the dipole moment. Since the infrared region's absorption is quantized, an organic compounds' molecule will have a number of peaks there. [3].

1.2.3. Near Infra-red Spectroscopy:

Near Infrared Spectroscopy (NIRS) is a very useful rapid, green, robust and non-destructive method. By using blackbend bulb thermometer and a glass prism, Sir Frederick William Herschel reported the first non-visible spectra in the year of 1800. As, this spectra lies on the below range from red light of visible spectra, so he named it "Infrared". In Greek language the word "Infra" means "below". Earlier the name of Infrared spectra was "calorific rays". NIR region is having short wavelength. NIR spectrum emission occurs when radiation energy related to the chemical bond of a particular molecule and the motion of the atoms transformed into the mechanical energy. NIR spectroscopic method commonly used today not only in the chemical, pharmaceutical, and food sectors but also in the forestry and agricultural sectors [10].

1.2.4. Applications of Near Infrared Spectroscopy:

Applications of near infrared spectroscopy can be segregate in different sectors,

a. Medical Application:

Oxygen saturation in hemoglobin is known as "sats", nowadays a technology has been developed using NIR spectroscopy. It can predict that level of saturation. In human body, to control the oxygenation and brains' micro-vascular system this technology has been used widely.

b. Agricultural Application:

NIR spectroscopy is commonly applied for determining the parameters of different agricultural commodities. This is the suitable method, as it is rapid, cost effective and non-destructive method. By using NIRS one can analyze the quality parameters of cereals, grains, dairy products, fruits, vegetables, tea, coffee, different types of beverages etc.

c. Material Science:

Measurement about the thickness of the thin films, optical properties of the nano-materials can be determined by the near infrared spectroscopic techniques.

d. Astronomical Application:

To study the environment of the cool stellar system NIR spectroscopy is a suitable method. As, different molecules can be form in the stellar system their vibrational and rotation spectra can be analyzed by near infrared spectra. The formation of molecular clouds also can be studied with the particular wavelengths of near infrared region. The dust particle of the interstellar medium can affect minutely by infrared wavelengths, this is known as reddening.

e. Remote Monitoring:

NIR hyperstructural imaging is very useful for structural determination of the organic compounds present in the several plants and soil. In the climate change and resolving many atmospheric issues also it has an impact. This kind of imaging is known as hyper-structural imaging, which can collect the data from NIR spectra.

f. To Measure The Size of Any Particle:

An important application of NIR spectroscopy is to measure small particles size related to agricultural grains and pharmaceutical grains [5].

1.3.Mango:-

Mango (*Mangifera indica*) is honored with the title of “king of fruits”, because of its great utility, delicious fleshy flavor occupied a pre-eminent place amongst the fruit crop grown in India. Also, it is having very high nutritious value containing with various micro and macro nutrients. In terms of production, acreage, and popularity, the mango is

one of the primary tropical and subtropical fruits. Numerous nutrients in mangoes promote good health. Mangoes, unlike many other fruits, are high in vitamin E. A (betacarotene), B6, Carbon, potassium, calcium, phosphorus, magnesium, copper, iron, zinc, and fiber are also included in mangoes. Additionally, mangoes are a good source of vitamins A, C, and B complex. Vitamin A is very good for human bones and eye sight and Vitamin C is good for immune system. Mangoes are having very good fiber content which can act like an ant carcinogenic element and helps to control the cholesterol and heart diseases [11].

1.3.1. History:

In Indian history, mango cultivation has a long history. From the documentation about mango it is known that 4000 years ago mango cultivation has started in ancient India. Although mango was very old cultivated fruit of ancient India but it was not noticeable to the world. So its demand was very less at that moment. But many foreigners came to this country for many purposes in different periods, and they have spread mango's popularity in all over the world. Like, Hiun Tsang (1st foreigner who took mango in notice to the world), IbnHaukal, IbanBatuta, Amir Khasru, AbulFazal etc. In the Mughal period, Mughal emperor Jalal uddin Akbar formed a mango garden in Delhi known as "Lakh Bagh", planted 1,00,000 mango plants [12].

Bengals' last "Nawab" SirajUd-Doula promoted this "king of fruit". In fact in his ruling period 124 varieties of mangoes planted over there. Among them many were cross breed and some are endangered too now. "Dilpasand", "Ranipasand", "Mirzapasand", "Kalasur" are mostly known among them [13].

1.3.2. Production:

The national fruit of India mango is cultivated through all over the India. It is a tropical fruit and cultivated not only in professional farms but also in common peoples' garden. As per the data obtained from National Horticulture Board, Government of India mangoes are cultivated in 2313 Hectare area of this country and its recent yearly production is 22353 Metric Ton (2018-19) [14].

1.3.3. Types of Mango:

The demands of worlds' tropical fruits are fulfilled by Mango. It is preferable because of richness in variety. The state wise types of available mangoes are tabulated below in Table 1.1.

Table 1. 1 State-wise distribution of Indian mangoes

States	Important Varieties
Andhra Pradesh	Baganpalli, Bangalora, Cherukurasam, Himayuddin, Suvanarekha, Tothapuri, Kesar, Dhasseri, Amrapali, Mallika, Neeleshan
Bihar	Bombai, Langra, Fazri, Himsagar, Kishenbhog, Sukul
Goa	Fernnandin, Mankurad, Alphanso
Gujrat	Alphanso, Kesar, Rajapuri, Vanraj
Karnataka	Kesar, Dasheri, Bangalora, Mulgoa, Raspuri Neelum, Totapuri, Baneshan, Alphanso, Mallika, Pairi
Kerala	Muthalamookkam, Mundappa, Pairi, Nadassala, Neelam, Suvarnarekha, Olour
Madhya Pradesh	Alphonso, Bombai, Langra
Maharashtra	Alphonso, Kesar, Mankurad, Mulgoa, Pairi, Rajapuri, Neelam, Totapuri
Orissa	Baneshan, Langra, Neelum, Suvarnarekha
Punjab	Dashehari, Chausa, Langra
Tamil Nadu	Immampasant, Bangalora, Rumani, Alphonso, Kalepad, Mulgoa, Banganpalli, Sendurga, Malguavo, Kallmai, Neelum,
Uttar Pradesh	Bombay Green, Dashehari, Fajri, Langra, SafedaLucknow, Chausa, Ratual, S.Saurabh, Amarpalli, Malihabadi, Bombay yellow
West Bengal	Bombai, Himsagar, KishanBhog, Langra, Bimli, Bharati, Champa, Dudhkumar, Khrishtapati, Rani, Rakhalbhog etc.

Source: [13]

Figure 1. 3 Some Indian Mangoes (Himsagar, Muloga, Totapuri, Badami, Bombay Green and Mallika) of Different Region
Source: [15]

1.3.4. Quality Parameters:

Quality is the ultimate benchmark of any product for its market value and quality depends on the different parameters. The different parameters are nutritional values, aroma and flavor and finally color and appearance. All four parameters are very important to judge about the mangoes, that it will accept by the consumers or not [16]. Quality of mango can be directly predicted by its pH value, dry matter content, acidity and sweetness (i.e. sugar content). Although internal quality of mangoes can't be determined by the consumers at the time of purchase, but the external parameters like, surface firmness, gloss, and aroma those can be judged easily [17].

1.3.4.1.Nutritional values

Mango is having rich quantity of various nutrients. Some are of them are macro nutrients (need very large amount for the body) and some are the micro nutrients (need very small amount for the body). The macronutrients present in mango are the carbohydrates, lipids, proteins, amino acid and different organic acids. And the micronutrients exists in mango are vitamins and different minerals.

a. Macronutrients

Macronutrients are the type of foods which needs in large amount in the diet and mangoes are having several macronutrients like carbohydrate, protein, fat, dietary fiber etc. An average estimation of different macronutrients has presented in Table 1.2.

Carbohydrate

Mango is having a good quantity of sugars and carbohydrate. Starch is present in the form of carbohydrate in the mango and pectin is structural carbohydrate. The main sugar content present in mango is glucose, fructose and sucrose. The sugar content and the carbohydrate content are responsible for the sweetness of the fruit. A fresh mango fruit flesh is having 15% sugar. Sugar content varies with the ripening stage of the fruit. Unripe mango fruit is having monosaccharide and sucrose is the main source of sugar when it turns into the ripe fruit. The mango fruit contains pectin, a structural carbohydrate that serves as the mango's sugar content. During the ripening stage, pectin hydrolysis occurs. Thus, pectin is a type of structural carbohydrate that accumulates in mangoes when they are young and green. As, the mango fruit ripens, the amount of pectin in the fruit decreases. Different cultivars have slightly different concentration of carbohydrate and sugar content [18].

Protein and different amino acids

Mango is not very good in the protein content. It differs slightly with the different varieties of mango. During the ripening stage the content of amino acids also varies. Glutamic acid, tyrosine, valine, lysine, leucine those are some amino acids present in mango. Glutamic acid is having the more proportion [18]. Although mango is not a large producer of protein, but research has been done to extract the protein content of mango pulp [19].

Lipid and different fatty acids

Mango pulp is the source of lipids (fat), although not in very sufficient quantity. It is an important component for the human body. But fatty acid cannot found within the mango flesh. The source of fatty acid is the mango peel and mango kernel. Fatty acid is the useful by-product one can utilize from the mango. These fatty acids can be utilized in

various food and pharmaceutical sector. Palmitic, stearic, oleic and linoleic are four main fatty acids present in mango kernel. And there many more fatty acids present in the mango seed and peel like linilenic, behenic acids, lignoceric etc., but with very less content. The fat content of mango seed can be act like a good replacement of cocoa butter [18]. Ripe mangoes are the excellent source of essential fatty acids and the amount of those fatty acids expands during harvesting period [20].

Table 1. 2 The macro nutrients present in the 100 g of mango pulp

Quality parameters	**Amount (in g)**
Water content	78.9-82.8
Ash content	0.34-0.52
Lipid content	0.30-0.53
Protein content	0.36-0.40
Carbohydrate content	16.20-17.18
Dietary fiber content	0.85-1.06
Energy (in Kcal)	62.1-190

Source: [18]

b. Micronutrients

Micronutrients are the foods which needs in very small quantity in human diet and it includes different types of minerals and vitamins.

Vitamins

There are several vitamins present in mango which makes nutritional value of the fruit richer. According to the USDA all mangoes are having two types of vitamins those are dissolved in water and fat. World Health Organization explained with their research that mangoes are having the large concentration of Vitamin A and Vitamin C. All vitamin contents present in mango pulp are having medicinal uses. During the ripening stage the vitamin content changes accordingly. The unripe mature green mangoes are having more amount of vitamin C, whereas the concentration of Vitamin C is less in ripe mangoes. Due to some metabolic pathways (i.e. tartrate biosynthesis, ethylene and the oxalate) the vitamin C content reduces in ripe mangoes. Retinol is an important vitamin within the

vitamin A group and one mango fruit can supply 15-69 retinol equivalents per day. The concentration of vitamin E and vitamin K is very less in mango fruits and content of vitamin E increases with the ripening stage within the mango fruit [18]. For some varities of mango vitamin E content varies directly with the ripening stage, but for some particular cultivars the vitamin E content varies inversely with the ripening stage. The reason of this is the formation of α-tocopherol, due to this the vitamin C content change and accordingly vitamin E content also varies [21-22]. Vitamin A content basically varies from 1000 IU to 6000 IU within all the varieties [23]. All vitamin concentration of mango fruit can vary with the varieties and the regions and it depends upon the pre and postharvest treatments [24]. An average estimation of different vitamin concentration has presented in Table 3.1.

Table 1. 3 The amount of different vitamin content present in 100 g of mango pulp

Vitamins	**Amount**
Vitamin C (Ascorbic acid)	13.2-92.8 mg
Thiamine	0.01-0.04 mg
Riboflavin	0.02-0.07 mg
Niacin	0.2-1.31 mg
Panthotenic acid	0.16-0.24 mg
Pyridoxin	0.05-0.16 mg
Vitamin A	54 µg
Vitamin K	4.2 µg
α-tocopherol	0.79-1.02 mg
Folate total	20-69 µg
Folate food	20-69 µg

Source: [18]

Minerals

According to the USDA research mangoes can produce essential minerals like Calcium (Ca), Iron (Fe), Magnesium (Mg), Potassium (K), Phosphorus (P) etc., which are very good for human health. Mangoes are having comparative good production of Potassium,

Phosphorus and Calcium while the concentration of Iron, Sodium and Zinc is very less in mango pulp [18]. Atomic absorption spectrophotometer is the best approach to detect the mineral concentrations [25]. An average estimation of different mineral contents has presented in Table 1.4.

Table 1. 4 The amount of different minerals present in 100 g of mango pulp

Minerals	**Amount**
Calcium	7-16 mg
Iron	0.09-0.41 mg
Magnesium	8-19 mg
Phosphorus	10-18 mg
Potassium	120-211 mg
Sodium	0-3 mg
Zinc	0.06-0.15 mg
Copper	0.04-0.32 mg
Manganese	0.03-0.12 mg
Selenium	0-0.6 mg

Source: [18]

1.3.4.2.Aroma and Flavor

1. Different Organic Acids

An individual's experience of food, water, or other substances is influenced by their organoleptic qualities, which include taste, sight, smell, and touch. So this property of the mango fruit along with the fruit quality and acidity has been controlled by the different organic acids. Those acids are also complimentary for the aerobic metabolism system. Fruit acidity can be measured by the pH value. And pH value records basically the presence of organic acids within the mango pulp. The concentration of organic acids may vary with the different mangoes [18]. As the parameters are synthesis of the acid, utilization and degradation, compartmentation and most important external effects (i.e. water supply, temperature, light etc.) during the ripening stage [26]. Although malic acid and the citric acid is the playing main role for the organic acids in mango, but still there are having some other organic acids with very less quantity like, oxalic acid, pyruvic

acid, galip acid, succinic acid, glucuronic acid, muonic acid, tartaric acid. So, for flavor organic acids keeps their main role [18]. The main source of organic acid is citric acid and malic acid in some Indonesian mangoes, but this variety has very tiny amount of ascorbic, tartaric and oxalic acid [27]. Some Indian varieties like Zardalu, Badami and Fazli have oxalic, citric, pyruvic, tartaric, succinic and malic acid content effectively [28-29], whereas, African mangoes have only the citric acid content [30].

2. Volatile Compounds

Aroma of the mango is responsible for the different volatile compounds in mango. Mangoes are having approximately 270 volatile compounds within it. Montoterpene and sesquiterpene are two major volatile components present in mango fruit. Apart from those two hydrocarbons lactones and esters have the major role for some varieties. Mangoes are not having any particular flavour but having a variation of taste. Due to variation of volatile compounds variety to variety the flavour changes accordingly. Odour activity value (OAV) is indicator which indicates the presence of volatile compounds in mango and is known as ratio of concentration of volatile compound in the food to its odour threshold. The aromatic foods are having more than 1 value of OAV. Volatile compounds can have glycoside structure and have very low molecular weight approximately less than 400 Da [31]. As volatile compounds have very high vapour pressure they can easily diffuse the aroma into the air, water and soil. Many methods can be used for those small amounts of organic compound extraction. Some of the useful extraction methods are liquid-liquid extraction, soil-phase extraction, static or dynamic headspace, Simultaneous Distillation Extraction and Micro Extraction in solid phase (SPME) [32].

Mangoes are having many volatile compounds responsible for its aroma but some of them are present with a good quantity like ethyl butanoate, (E,Z)-2,6-nonadienal, ethyl-2-methylpropanoate, (E)-β-ionone, methyl benzoate, decanal and 2,5-dimethyl-4-methoxy3(2H)-furanone, (E)-2-nonenal [31].

1.3.4.3.Color and Appearance

Color and appearance is an external attribute of the mango fruit. It changes during the

ripening process of the fruit. This color component is dependent on the natural pigments present in the fruits [16]. Those natural pigments can be chlorophylls, flavonoids and carotenoids [18]. Also the maturity of a mango fruit is being indicated by the different factors i.e., skin color, shape, size and shoulder growth [33]. The main reason of color changing ripening and maturation stage is the degradation of chlorophyll pigment and increase the percentage of polyphenols or carotenoids.

Chlorophylls

The concentration of the chlorophyll pigment varies with the several cultivars of the mango. The hard green immature mangoes are having green peel color and presence of the chlorophyll pigment is responsible for that. Color is important external parameter which reflects the fruit quality initially and the maturity of that particular fruit. As the ripening day goes the green color changes into orange, yellow or red. Two types of chlorophyll present in mango that is chlorophyll a and chlorophyll b. Chlorophyll a is responsible for the blue green color of mango peel and the another chlorophyll b is the source of yellow green colour. Research predicted that these two types of chlorophyll present in mango with 3:1 concentration [18]**.** Chlorophyll concentration within the mango fruit reduce with ripening stage. The mature green unripe mangoes are having more chlorophyll content as compare to yellow ripe mangoes. The presence of β-carotene varies inversely with the Chlorophyll concentration within the mango peel [34].

Carotenoids and Flavonoids

Mango fruits are the sufficient source of carotenoids and this is fat-soluble compound. The major cultivars of mangoes are having the orange-yellowish skin and pulp colour due to the presence of carotenoids. Some of ripe mangoes are red in color. Not only carotenoid but also at some point the compound anthocyanin has the impact for this reddish color. Carotenoids work as the accessory pigment (Those pigments are light absorbing compounds which found in photosynthesis organisms) and as an antioxidant substance. This compound placed at chloroplasts in the chlorophyll tissues. There are many carotenoids present in the mango such as α, β and γ-carotene. The carotenoid concentration in mango depends upon the harvesting process and the environmental

ambience. There are some reasons for the concentration variation of carotenoids like state of maturation, genetic factor, postharvest handling and production techniques etc. [18]. Among all different carotenoid present in ripe mango pulp β-carotene produces the 60% of the essentiality [19], but for the green mango has majorly produced the lutein [35].

Flavonoids are the most important phytochemical (polyphenols) present in the mango fruit. There are many major flavonoids like flavonols, flavanones, flavones, acthocyanidins and flavan-3-ols. Among the all major dietary polyphenols (stilbenes, lignanes and flavonoids) flavonoids can produce the maximum source (near about 60%) [18].

Size and Sphericity

Appearance of the mango fruit can be characterized with one more physical parameter that is size and sphericity. These are the external attributes of a mango. Size can be measured by a vernier caliper apparatus and determined with a formula of geometrical mean diameter. That is,

$$\text{Size} = (abc)^{1/3} \tag{7}$$

Where, “a” is the intercept of the length, “b” is the normal to that intercept “a” and “c” is the longest intercept normal to “a” and “b” parameter and sphericity of a mango fruit can be determined with the formula as [36].

$$\text{Sphericity} = \frac{\text{Geometrical mean diameter}}{\text{Longest intercept}} \tag{8}$$

Longest intercept can be ‘a’, ‘b’ or ‘c’.

1.3.4.4.Rheological Properties:

Mango has occupied a large position within the world market. Among the all kind of commercial fruits, mango is considered in almost top of them. For the industrial and technological application it is essential to have a proper knowledge about the rheological properties of the mango. Not only intact mango fruits but also some other byproducts of mango fruit like pulp, puree, leather etc. can have proper usage. Among them pulp is having a great acceptance in the food processing industry. One of the most essential factors, the rheological property is helpful in projects, evaluations, and the operation of

process equipment including agitators, pumps, heat exchangers, and separations in addition to serving as a quality indicator [14, 37-42].

Mango pulp is a viscous liquid and the flow characteristics of this liquid can be the result of rheological findings. Mango pulp is a pseudo plastic liquid and this property of the mango pulp can be proved with the coaxial cylindrical rheometer. Rheometer can predict the properties are time dependent or time independent. Mango pulp is also having the thixotropic property and is having high yield stress [37, 40].

1.4.Multivariate Analysis

By examining the data covariance structure, multivariate analysis is employed to comprehend and reduce the data dimensions. The direction of the linear relationship between two continuous variables is now determined by the covariance. Multivariate analysis reveals the data's structure and significance through the use and interpretation of additional statistical techniques not covered in a basic statistics course. In addition to focusing on the numerous variables present in the dataset, multivariate analysis also seeks to establish relationships between groupings of variables or divide entities into various subgroups. A larger dataset with more variables is the focus of multivariate analysis [43, 44].

1.4.1. Motivation of the Multivariate Analysis

There are several motivation are followed to do the multivariate analysis,

- **Data Reduction:** Identify one or a few variables that capture most of the variability in a large dataset.
- **Sorting and Grouping:** Develop rules for splitting a large dataset into relatively homogeneous groups.
- **Dependence:** Characterize relationships among variables and use of dependence variable to predict.
- **Hypothesis Testing:** Multiple responses in a designed experiment [45].

1.4.2. Use of Multivariate Analysis

There are several important uses of the Multivariate analysis. The uses of multivariate analysis are followed,

- **Complexity**

It's possible that the topic or data under study is more complex than what can be analysed using univariate approaches.

- **Reality**

In some circumstances, conducting a univariate analysis is not appropriate because the facts or research require a multivariate analysis.

- **Experimental vs. Empirical data**

Experimental data generally manipulates a single variable making univariate analysis easier to imply some causality; empirical data however may require multivariate analysis for accurate correlation analysis.

- **History**

Multivariate analysis calculations were much harder due to lack of availability of statistical computing, advances in computing and software makes it much easier to perform [45-47].

1.4.3. Caution for Multivariate Analysis

- **Ambiguity**

Multivariate analysis is a less clear understanding of the data. The example can be, the group differences on a linear combination of dependent variables. It is sometimes easier to explain group differences in a univariate context.

- **Complexity**

Just because a technique is popular does not mean anyone needs to use it.

- **Assumption**

Multivariate analyses have rules and assumptions like univariate analysis [46].

1.4.4. Advantages and Disadvantages of Multivariate Analysis

a. Richer, more realistic design.
b. Examines events from a broad perspective.
c. Each method differs in amount.
d. Can help control for Type-1 error (denying the null hypothesis even though it is true).

e. Numerous samples are frequently needed.
f. More challenging to interpret
g. Less known about the robustness of assumptions [48].

1.4.5. Different Approaches

Multivariate analysis or chemometrics is one of the most important statistical techniques which lead the regression methods like partial least square regression and principal component regression modeling. This method has extracted the data from the large dataset and thus it is useful for quality analysis of food or for analytical chemistry. Multivariate analysis can be considered as more preferable strategy than univariate analysis, because it can be interpret comparatively large dataset. This method is very useful for qualitative and quantitative analysis of any material. Regression modeling (multivariate analysis) is generally required for predicting the quality parameters with a green and nondestructive way. Multivariate analysis can identify, quantify and estimate the quality of the sample with the association of several regression methods like PLS and PCR [49].

- **Principal Component Regression**

Principal Component Analysis is used to create linear models from large data sets (PCA). Orthogonal vectors (loadings) are used in PCA models to build the model. The purpose of PCA is to reduce the size of complicated data sets, avoid noise associated with undesirable or outlier samples or variables, and limit the impact of random error. PCA can explain the covariance of data by employing a new set of variables called principle components (PCs). We can create relationships between samples (scores) and variables by using PCs, which are linear combinations of original data (loadings). PCA is the first step in the analysis of the large data set, followed by clustering/classification and additional chemometrics methods such as PCR. The anticipated data set's maximum variance is indicated by the first PC. The second PC is orthogonal to the first and accounts for the majority of the variance among the projected data points. In this manner, various PCs are used to cover the entire expected data set. The identification of

correlations between samples and variables as well as the identification of the outlier in the sample data sets are two benefits of data reduction based on PCA [50].

- **Partial Least Square Regression**

PLSR method is important to identify a coordinate system that is preferable to the one used by PCA and PCR. In that case, PLS regression goes one step further than PCA. Both spectral and reference data (analytical data) are impactful in PLS and with the analysis of both the data several PLS factors are generated. It first employs PCA's capacity to reduce the noise and errors of the reference and spectral data. To produce a noise-free spectrum, PLS spins the vectors until they are aligned. Additionally, new coordinate axes are obtained for the reference data (y variable) and spectral data (x variables). Calibration, validation, and prediction are the three stages of the PLS method. Using the spectrum and the reference data, a regression model is created for calibration and validation. Root mean square error of calibration (RMSEC), root mean square error of validation (RMSEV), and coefficient of correlation all contribute to the model's correctness (R_c^2 and R_v^2). Using the new samples' spectral data, the model is then used to forecast their parameters [51].

- **interval PLS (i-PLS)**

This is basically a wavelength selection method used for spectroscopic data for qualitative and quantitative analysis. Main idea of this method is to apply regression models on different parts of the data and at specific intervals to obtain the most relevant wavelengths required for prediction of a particular parameter. i-PLS is employed in two different steps

- The complete spectral range is split in N intervals of equal width.
- Regression models are applied in forward direction on the intervals starting from lowest to highest wavelengths. The last step is to compare the regression models of all the intervals and neglecting the models with less value of correlation coefficient or high values of root mean square error [52].

- **Synergic Interval PLS (si-PLS)**

This regression method is used the combination of differently equally placed intervals together to develop a regression model. si-PLS is employed to find the robust modes with

less error and with more accuracy, using least number of important wavelengths. The regression model with lowest root mean square error is selected for prediction of that particular parameter [53].

Use of i-PLS and si-PLS methods for regression models helps in development of rapid and robust model. These techniques have been used by many authors to create regression method with selected wavelengths resulting in low cost, rapid and robust models [5].

References:

1. R. Kumar, *Atomic and Molecular Spectra: Laser* (Kedar Nath Ram Nath, Thoroughly revised edition, 2008, Reprint 2018).
2. C. Wilson and G. Mcgranaghan, *Reinf. Plast.*, **58**, 43-47 (2014).
3. Y. R. Sharma, *Elementary organic spectroscopy* (S Chand, Fifth edition (2013)).
4. D. J. Pietrzyk and C. W. Prank, *Analytical Chemistry* (New York: Academic Press, 1979), pp. 450-465.
5. D. A. Burns and E. W. Ciurczak, *Handbook of Near-Infrared Analysis (Practical Spectroscopy* (CRC Press, 3rd edition, 2007).
6. P. L. Kalsi, *Spectroscopy of Organic Compounds* (New Age International Publishers, Seventh edition, 2016).
7. L. M. Ng and R. Simmons, *Analyt. Chem.*, **71**, 343-350 (1999).
8. N. C. Thomas, *J. Chem. Educ.* **68,** 631 (1991).
9. M. R. Derrick, S. Dusan, and M. L. James, *Infrared spectroscopy in conservation science.* (Getty Publications, 2000).
10. C. Pasquini, *J. Braz. Chem.* Soc., **14**, 198-219 (2003).
11. G. Reich, *Adv. Drug Deliv. Rev.*, **57**, 1109-1143 (2005).
12. Official website of Research and Development, Indian Institute of Technology Kanpur, "Fourier Transform Infrared Spectrometer", (2012), Official link to be followed: https://www.iitk.ac.in/dordold/index.php?option=com_content&view=category&layout=blog&id=221&Itemid=240 (accessed on 2nd February, 2022).
13. Post-Harvest Profile of Mango, Govt. Of India, Ministry of Agriculture (Department of Agriculture and Cooperation), Directorate of Marketing and Inspection, Brach Head Office Nagpur (2013).
14. S. N. Jha, K. Narsaiah, A. D. Sharma, M. Singh, S. Bansal, R. Kumar, *J. Food Sci. Technol.*, **47**, 1-14 (2010).
15. M. De., J. Das, M. Saha, A. Pal and S.R. Dey, *J. Environ. Sociobiol*., **11**, 187-198 (2014).

16. C. Zhen, P. Ong, H. Norhashila and M. Bernard, *J. Agric. Food Engineer*., **1**, 0003 (2020).
17. S. N. Jha, J. Pranita, K. Narsaiah, M. Gupta, R. Bhardwaj and A. K. Singh, *Sci. Hortic*.,**138**, 171-175 (2012).
18. C. Maldonado, E. Maria, M. Y. Elhadi, B. Ramiro, L. Patricia, L. Nelsy, A. Johanny, R. Beatriz, and C. G. O. Juan, *Front. Plant Sci*., 10, (2019).
19. M. S. Dar, P. Oak, H. Chidley, A. Deshpande, A. Giri, and V. Gupta, *Nutritional Composition of Fruit Cultivars* (Academic Press, 2016), pp. 445-467.
20. A. B. Deshpande, H. G. Chidley, P. S. Oak, K. H. Pujari, A. P. Giri and V. S. Gupta, *Data Br*., **9**, 480-491 (2016).
21. G. I. Barbosa, , K. P. Caballero Montoya, N. Ledesma, S. G. Sáyago Ayerdi, M. D. L. García Magaña, , E. J. Bishop von Wettberg and E. Montalvo-González, *J. Sci. Food Agric.*, **97**, 4987-4994 (2017).
22. R. K. Singh, S. A. Ali, P. Nath and V. A. Sane, *J. Exp. Bot*., **62**, 3375-3385 (2011).
23. A. C. Matheyambath, J. Subramanian and G. Paliyath, *Encyclopedia of Food and Health*, (Ed. B.C., Paul, F.F. Told, (Switzerland: Elsevier), 2016), pp. 641-645.
24. J. Aslam, A. Mujib, Z. Fatima and M. P. Sharma, *In Vitro Cell. Dev. Biol.*, ***46***, 348-353 (2010).
25. M. Bala and J. B. Bashar, *Bayero J. Pure Appl. Sci*., 10, 172-176 (2017).
26. J. G. Vallarino and S. Osorio, *Postharvest Physiology and Biochemistry of Fruits and Vegetables*, (Woodhead Publishing, 2019), pp. 207-224.
27. A. P. Medicott, A. K. Thomson and H. J. Lalel, *J. Sci. Food Agric*., **36**, 561-66 (1986).
28. M. S. Shashirekha and M. V. Patwardhan, LWT *Lebensmitt Wissensch Technol*, (1976).
29. S. Kumar, D. K. Das, A. K. Singh and U. S. Prasad, *Indian J. Plant Physiol*., **36**, 107-111 (1993).
30. O. C. Othman and G. P. Mbogo, *Tanz. J. Sci*., **35**, 57-65 (2009).

31. J. A. Pino and M. Judith, *Flavour Fragr. J*., **21**, 207-213 (2006).
32. J. A. Pino, J. Mesa, Y. Muñoz, M. P. Martí, R. Marbot, *J. Agric. Food Chem*., **53**, 2213-2223 (2005).
33. S. N. Jha, S. Chopra and A. R. P. Kingsly, *J. Food Engineer*., **78**, 22-26 (2007).
34. S. Ketsa, P. Wandee and S. Suranant, *J. Plant Physiol*., **154**, 363-366 (1999).
35. M. Kato and S. Shimizu, *Plant Cell Physiol*., **26**, 1291-1301, (1985).
36. N. N. Mohsenin, *Thermal Properties of Food and Agriculture Materials* (Gordon and Breach Science, New York, 1980), pp. 407.
37. D. H. Pelegrine, F. C. Silva and C.A. Gasparetto, LWT-*Food Sci. Technol*., **35**, 645-648 (2002).
38. M. Siddiq, J. K. Brecht, and J. S. Sidhu, *Handbook of Mango Fruit: Production, Postharvest Science, Processing Technology and Nutrition*, (First edition, John Wiley & Sons., 2017), pp. 1-16.
39. M. Dak, R. C. Verma, and S. N. A. Jaaffrey, *J. Food Engineer*., **80**, 1011-1015 (2007).
40. H. A. El-Mansy, A. M. Sharoba, H. E. L. M. Bahlol and A. I. El-Desouky, *Ann. of Agric. Sc., Moshtohor*, **43**, 665-686 (2005).
41. B. Manohar, P. Ramakrishna and R. S. Ramteke, *J. Texture Stud*., **21**, 179-190 (1990).
42. V. E. Nambi, K. Thangavel, K. A. Rajeswari, A. Manickavasagan and V. Geetha, *Postharvest Biol. Technol.*, **117**, 152-160 (2016).
43. M. S. Bartlett, *Suppl. J. R. Stat. Soc*. **9**, 176-197 (1947).
44. N. Martin and H. Maes, *Multivariate Analysis* (Academic Press, London, 1979).
45. H. Abdi, *Multivariate analysis. Encyclopedia for research methods for the social sciences*. (Thousand Oaks: Sage, 2003), pp. 699-702.
46. H. Martens and M. Martens, *Multivariate Analysis of Quality: an Introduction*. (John Wiley & Sons., 2001)
47. C. Chatfield and A. J. Collins, *Introduction to Multivariate Analysis*, (Routledge, 2018).

48. J. D. Finn, *A General Model for Multi-variate Analysis*, (Holt, Rinehart, Winston. Inc., New York, 1974).
49. Møller, S. F., von Frese, J., & Bro, R. (2005). *J. Chemom.*, *19*(10), 549-563.
50. Næs, T., & Martens, H., *J. Chemom.*, **2**(2), 155-167, (1988).
51. Mehmood, T., Liland, K. H., Snipen, L., & Sæbø, S., *Chemometr. Intell. Lab. Sys.*, **118**, 62-69, (2012).
52. Norgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B., *Appl. Spectros.*, **54**(3), 413-419, (2000).
53. Deng, B. C., Yun, Y. H., Ma, P., Lin, C. C., Ren, D. B., & Liang, Y. Z., *Analyst.*, **140**(6), 1876-1885, (2015).

Chapter-2: Literature Review

Chapter-2

Literature Review

A systematic process for locating, assessing, and interpreting research work generated by academics and researchers is known as a literature review. It can also be viewed as a crucial task for educating readers about the body of information that is currently available in a given field of interest. This chapter focused on the literature survey to identify the research gap of the concerned area.

Saranwong et al. (2001) have compared two different regression method, one is multiple linear regression and another one is partial least square regression. Both the method have proved as efficient for similar results for estimation of dry matter content and total soluble solid (in $^{\circ}$Brix) content with a time effective method. They have taken two several near infrared regions (short wave (700-1100 nm) and long wave (1100-2500 nm)) and results indicated that short wave region were more promising. The results were improved using some selective wavelengths from the total range. They have used multiple linear regression to select those particular wavelengths [1].

Saranwong et al. (2003) have discussed about the eating quality of ripe mango, which is impactful for the economical existence of the fruit. For this purpose they have checked and chosen the hard green mango before the ripening. They have finally determined dry matter content and starch quantity with near infrared spectral imaging. Both the parameters are required for maintaining the eating quality of hard green immature mangoes before market use. They have also studied the prediction of the fruit quality during the harvest period [2].

Saranwong et al (2003) have compared two different devices, technically different

lab made portable near infrared spectrometer and laboratory NIR spectrophotometer. They have compared the usefulness of these two technologies and evaluated the impact of result using the regression analysis. They have estimated the ˚brix value of the mango pulp with portable FT20 NIR spectrometer and NIR FOSS6500 to determine the same. They have obtained for both the cases the prediction was similar and hence proved the same efficiency of both instrument [3].

Saranwong et al. (2004) have discussed about a rapid and non-destructive method to ensure the different quality parameters of ripe mangoes. Maturity indices (flesh color and flesh carotenoids) are the indicator to check the harvesting quality of the green unripe mangoes. They have taken 288 samples of "Mahajanaka" mango from cultivation farms of Lumpang, Thalinad. They have kept as it is (natural hard green mango) and some of them were kept in a conditional temperature of 25^{o}C with an artificial stimulator for mango i.e. Calcium carbide. They have discussed about the eating quality of the mangoes which were kept at 25^{o}C temperature with flesh firmness, soluble solids content and titratable acidity parameters. They also have determined physico-chemical parameters like density and the specific gravity peel and flesh colour, total sugar content, starch content and dry matter, total sugar content (glucose, sucrose and fructose) content. Initially they have observed that at 105^{th} day after fruit set total sugar content was 3.28 but at 140^{th} day after fruit set it were 3.52. Specific gravity also was 1.005 initially but finally it was 1.014 after 140 days of fruit set. They have obtained SSC, TA and firmness as16.64, 0.37 and 7.75 for 140 day after fruit set, 133 day after fruit set and for 140 days after fruit set respectively. Now by using non-destructive (NIR) approach also they have studied the eating as well as harvesting quality of two different mangoes (mature and immature). For mature mangoes the prediction of harvesting quality with NIR calibration equation the value of dry matter content reported as 18.54 and for the mature mangoes the same reported as 21.15. Starch content for immature mangoes is 39.4 and for mature mangoes it is 46.76 [4].

Mahayothee et al. (2004) have discussed the usefulness of the near infrared spectroscopy for determination of weight loss, fruit firmness, total soluble solid and titratable acidity. They have developed a regression method with the NIR spectral data and this model was developed to determine the above mentioned properties non-destructively. Nam Dokami and Chok Anan these two particular Thai cultivars, they have taken for the analysis. For every parameter, they have obtained the coefficient of determination more than 0.9, which had indicated that NIR is an applicable tool for prediction of the quality of mango fruit during the ripening stage [5].

Walsh et al. (2004) have discussed about the chemmometric approach to resolve the grading and sorting of the fruits and vegetables. They have used portable near infrared spectrometer made with tungsten bulb, detectors and transmittance optical fiber to generate the result. They have resolved a real life quality assurance challenge with fruit sample mango and banana and as vegetable samples potato and tomato. The analysis had been done after determination of root mean square values, coefficient of regression values and standard deviation of regression values. They have developed regression model to estimate the soluble solid content. The skin thickness was obtained for grading and sorting of the concerned fruit and vegetables [6].

Ahmend et al. (2005) have discussed the about the two different biochemical properties like pH and total soluble solution of two different mangoes that is alphanso and chausa. They have analyzed the change of those parameters after applying high pressure. For determining the several rheological parameters like yield stress, consistency coefficient and flow behaviour index, they have implemented different flow model like Casson, power law, Bingham, Newtonian and Herschel-Bulkley model. Among them Herschel-Bulkley model proved best as the standard error was least [7].

Subedi et al. (2007) have produced a regression statistical model to estimate the dry

matter content and total soluble solid of mango fruit. They have developed regression model using combined technology of the short wave near infrared spectroscopy (750-1100 nm) and chemometric approach. Also the correlation analysis between the dry matter content and flesh color has been performed and correlation has determined as 0.9079 [8].

Delwiche et al. (2008) have studied about the soluble solid concentration and the individual as well as combined concentration of all sugars (glucose, sucrose, and fructose) of mango. Regression equation using multivariate analysis (partial least square regression) and near infrared spectral data (range from750 to 1100 nm) has been developed. The method produced a rapid and green method to estimate the same properties. For all the cases the value of regression coefficient varied from 0.70 to 0.80. The study determined the application of near infrared spectroscopy for a nondestructive evaluation [9].

Valente et al. (2009) have discussed about two different methods these are Vis/NIR spectroscopy and another is acoustics impulse technique to estimate the firmness of mango with non-destructive way. During climacteric ripening, it is important to maintain the firmness in a fruit (here it is mango) for export and trading purpose. The term “shelf life” and eating quality of a mango is very significant and it is related to firmness of fruit. They have collected 145 samples of Kent mangoes and 171 samples of Keitt mangoes from the market of Senegal, South Africa and Andalusia, Spain respectively. They have used penetrometer to determine the results of different firmness factors. Mean value for firmness was determined of Kent and Keitt mangoes as 16.86 and 17.49 N-mm respectively and finally they have used the data to calibrate a model (F_P model) to calculate the firmness of ripe and unripe mango non-destructively [10].

Yu and He (2009) have obtained the rapid determination of soluble solid content in

mango fruit. They have determined the results using visible and near infrared spectroscopy. They have used two different algorithms like Back-propagation artificial neural network (BP-ANN) and partial least square (PLS) regression together for the estimation of soluble solid content. They have determined the error of prediction and coefficient of determination as 0.865 and 0.757 respectively. Their research also concluded that the results are more accurate and efficient in application of PLS and BPANN together. In this case, they have predicted the results better only with PLS regression analysis [11].

Subedi et al. (2009) have discussed about two technical approaches for determination ripening index during the harvest period. For firmness determination, they have used penetrometer for taking the reference values and after this they have developed a calibration model with spectral range of 400-1100 nm. Also they have taken the sound velocity method for collecting the reference values. For both the cases, they have analyzed the maturity of three fruit i.e., banana, mango and peach. They have performed correlation analysis between the data of sound velocity of all three fruits and fruit firmness and they have obtained that a good correlation with the correlation coefficient of 0.8 between these two parameters in case of mango, whereas the correlation was very poor in case of peach and banana. For these fruits, the short wave near infrared spectroscopy is not efficient for accurate results for prediction of fruit maturity, whereas sound velocity was the useful tool for the analysis of same [12].

Nagle et al. (2010) have been determined the impact of irrigation on the maturity index determination of mango with the optical measurements (NIR Spectroscopy). They have taken the samples from irrigated and non-irrigated orchards at different harvest time. They have obtained the NIR spectra for both (irrigated and non-irrigated) the samples and it was observed the difference in spectra. They have obtained a high positive correlation with the NIR determinations and dry matter content of irrigated samples with 0.80, correlation coefficient and a poor correlation was obtained with the

NIR determinations and dry matter content of non-irrigated samples with 0.57, correlation coefficient [13].

Saranwong et al. (2010) have investigated effect of fruit fly eggs and larvae during the harvesting period of mango. For this research, authors have imposed externally fruit fly eggs and larvae into the hard green matured mangoes. The spectra for both the fresh and affected mangoes had been taken and the analysis was done with two different spectral ranges. Long wave near infrared region is less effective than the short wave near infrared region (700-1100 nm). The study with partial least square discriminant analysis has been obtained an extra peak at 730 nm for the affected mangoes. So, that particular peak had indicated the presence of fruit fly egg and larvae and the study was developed a strategy for the future analysis of this kind of fruit damage parameters with a non-destructive and rapid manner [14].

Gundurao et al. (2011) have studied the thermal properties of mango pulp with the variation of total soluble solid content and temperature. In thermal properties, they have discussed about heat capacity, thermal diffusivity, density and thermal conductivity. In rheological parameter they have obtained shear stress, shear rate, consistency coefficient and flow behavior index with different standard model (i.e., Casson, Bingham, power law and Herschel-Bulkley model). And among these models power model has proved best for this case as it had standard error. They have determined the flow behavior index of mango pulp for different concentration of TSS and temperature using the power law model [15].

The NIR imaging technology to discriminate the quality of fresh mangoes and fruit fly affected mangoes have been discussed by **Saranwong et al. (2011)**. They have chosen a spectral image from the 400 to 1100 nm range of affected mangoes and fresh mangoes. They have utilized iterative Bayesian discriminant analysis to select the important wavelengths for the prediction fruit quality before and after of affected with

fruit fly. The concerned technology has used efficiently as cost effective tool for mango during harvest [16].

Two intrinsic components, pH and Total soluble solids (i.e. TSS) have been determined by **Jha et al. (2012)**. It is essential to maintain sweetness of mango and predict the intrinsic properties through any suitable, eco-friendly approach for its commercial value. Near Infrared spectroscopy is the one by which one can analyze the accurate value of intrinsic properties. Seven different types of mangoes have been taken from the seven different states of India as a sample. Chausa, Dasheri, Malda, Neelum, Kesar, Langra, Mallika took as a samples. Those samples kept into storage in a conditional temperature and moisture. With the digital refractometer and pH meter conventionally TSS and pH value determined but non-destructively by analyzing the NIR spectra a calibration and validation model has been made. Mean value of TSS and pH has reported as 14.9 and 3.81 respectively after calibration and validation. Data has been analyzed by the PLS and MLR regression model, statistically [17].

Some usual non-destructive approaches have been introduced by **Kitthawee et al. (2012)** to study the maturity and ripeness of the "King of fruits". They have done a comparative analysis of the properties of mango by three different methods that are electronic nose, visible light spectrometer and Near Infrared spectrometer. To check the maturity and ripeness, unripe green hard "Tommy Atkins" mangoes imported from Spain and Mexico has been collected from UC Davis campus. They have used both destructive and non-destructive method to predict the properties. They used three different approaches namely NIR spectra, visible spectra and electronic nose to predict the value of DM and SSC, flesh colour, volatile colour respectively. So, basically the ability of those techniques was being checked by them for analyzing the quality of mango. They have conclude that immature and mature 'Tommy Atkins' mangoes are having 70% and 95% dry matter and soluble solid content respectively. Finally, they have established an optical model for non-destructive predictions of those properties

[18].

Near infrared spectrometer has been designed by **Izeneid et al. (2012)** for assured the quality parameters of unripe green mangoes. Mangoes are very rich fruit with number of nutritious values. To predict its quality is very essential for its commercial value. For instrumentation voltage supplier, a microcontroller, an amplifier and optical sensor as switch needed. An optical sensor based spectrometer has been developed for non-destructive calculation. Technically, they have defined every part of a spectrometer beautifully and also a development of software of computational analysis with mikroC programming language has been done. They have designed a movable NIR spectrometer for analyzing the data with functional wavelength of 940 nm. For both the cases of ripe and unripe mangoes, they have taken reading and analyzed the impact of temperature and the pressure with sample respectively. As a further scope this work gave an idea about sugar content and other properties by their own designed instrumentation [19]. This same work has been published in other proceedings and the bio optical approaches has been done by **Izneid et al. (2013)** for predicting the quality parameters non-destructively [20].

Haff et al. (2013) have discussed the effectiveness of NIR hyperstructural imaging techniques to identify the affected mango fruits by *Bactrocera dorsalis* (fruit fly). The method has developed a way to grade and sort the mangoes after harvesting, before going to the market. This study resulted a way to produce quality fruits to the market for the consumers. Total sample size was divided into two parts; one segment was naturally ripened with proper care and in other segment fruit fly larvae was artificially implemented. Gray scale image was produced and some algorithm was used to reduce the overlapping clusters (particle size analysis), to eliminate the drift from image background (background subtraction), to reduce the noise (Gaussian blur). In terms of error rate determination from these NIR imaging the method proved its reliability [21].

Watanawan et al. (2013) have discussed about the quality indices of a specific mango cultivar that is "Nam Dok Mai". This study was focused on the near infrared spectroscopic study to determine the flesh colour, peel colour, dry matter content, total soluble solid (TSS) content and titratable acidity (TA). These parameters had an impact on fruit maturity. A harvesting stage from84 to 119 days were taken for analysis during the fruit ripening. The ripening was directly proportional to the dry matter content and TSS values. NIR values were taken in terms of R^2 and RMSE for calibration and validation both. The present study indicated that the utmost results during the 105-112 days of ripening [22].

Rivera et al. (2014) have elaborated a real life industrial and social problem and has applied the technology hyperstructural imaging to resolve it. Nowadays, to maintain market value it is very important to maintain the external quality of mangoes. Visually the quality cannot be checked properly and the damages for postharvest handling from farm to market always have affected the quality and have decreased the consumer satisfaction. To resolve this problem, they have utilized the application of NIR hyperstructural imaging technique. In this method, they have selected the spectral range from 650-1100 nm for analysis. They also have used five different classifiers (linear discriminant analysis, k-nearest neighbours, extreme learning machines and decision trees) to divide the samples into two parts that is damaged and not damaged [23].

Theanjumpol et al. (2014) have been reported a research about the screening capability of NIR spectroscopic technology to predict the physicochemical attributes of mango. In physicochemical properties, fruit firmness, skin colour, titratable acidity, total soluble solid and dry matter content has been calculated. Samples kept with different ambient conditions (humidity, temperature and harvesting stage) have been studied for analysis of the physicochemical properties. The estimation of these properties was done with PLS regression method. This method developed a regression equation and determined the

values of coefficients and errors. The study has concluded with the efficiency of the NIR technology in food quality analysis [24].

Impact of microwave vacuum drying method on the mango slice for the water dehydration process has been discussed by **Pu et al. (2015)**. A regression model was developed with spectral data and reference moisture values. They have taken the spectral data into two parts (vis-NIR spectra with 400-1000 nm range and NIR range with 1000-2500 nm range). Partial least square regression (PLS) was applied to develop a calibration model, which is the widely used method for much research. They have applied stepwise regression and competitive adaptive reweighted sampling strategies to select the important wavelengths for model development. The selected wavelengths are 908 nm, 1076 nm, 1153 nm, 1405 nm and 1706 nm has played an important role for prediction of moisture during microwave vacuum drying method. They have proved NIR spectroscopy as a promising tool to estimate the same with a time and cost effective way [25].

Different conventional methods (like titration) for measuring the expected quality parameters have been used by **Munawar et al. (2016)**. Some chemical properties of mangoes like total sugar content, ascorbic acid, titratable acidity can be predict by non-destructively with the infrared spectrometer. They have taken “Kent” mangoes as a sample from the local market of Gottingen, Germany. With the collected NIR data and determined SSC and TA of mango pulp and juice, they have developed calibration and validation regression model and that is known as chemometric. Due to calibration and validation the mean value of SSC has reported as 13.39, mean value of ascorbic acid has reported as 32.21 and titratable acidity has reported as 495.48. They have taken principle component analysis for classification of the samples. By using FT-NIR after determining all the parameters properly ensured the quality of an intact mango [26].

Sugar content, texture properties and rheological properties of three different Indian mangoes that are alphanso, baganapalli and neelam have been discussed by **Nambi et al. (2016)**. They have studied the texture properties for mango pulp by analyzing some parameters like peel strength, stiffness and flesh firmness. They have utilized different kinetic models and shear stress and viscosity were determined as rheological parameters. Finally, they have applied the suitable Herschel-Bulkley model to study the yield stress value [27].

Pu et al. (2016) have estimated the moisture content present in mango with four different slices. The effect of different shape of mango slice was studied. They have decided to take four different geometrical shapes of mango slices i.e., round, square, rectangular and square rectangular into the consideration for the moisture analysis. They have taken the micro oven dried mango slices and the near infrared spectra were recorded. They have applied Partial Least Square regression to select some particular wavelengths from total range (700-2500 nm). The selective wavelengths for accurate prediction of mango slices are 951 nm, 977 nm, 1138 nm, 1362 nm, 1386 nm, 1420 nm and 1440 nm. Micro wave oven drying was proved as an efficient strategy for determination of round shape mango slice with the better results. They have concluded the results that evaluated in terms of root mean square error and coefficient of determination values. Basically, they have analyzed the effect of microwave drying method for geometrical shapes of mango by using near infrared hyperstructural imaging [28].

Rungpichayapichet et al. (2016) have discussed the usefulness of the near infrared spectroscopy as a quick technique to measure the ripening index of the mango fruit. For this study, they have considered the short wave near infrared region (700-1100 nm) and performed second order derivative to modify the spectra. They have collected the reference data with continuous three postharvest periods. With one year data the model had given the low accuracy, but the combined data of the three years had given

the best model for prediction of ripening index. They have concluded the impact of the postharvest time period with the near infrared spectral data [29].

A rapid quality assessment method for Tommy Atkin mangoes in association with near infrared spectroscopy has been demonstrated by **Marques et al. (2016)**. They have determined several quality parameters namely firmness, soluble solid content, dry matter and titratable acidity. Standard normal variate preprocessing strategy has been utilized to modify the spectral data by them. They have obtained the results with particular wavelengths more accurately. They have used Jack Knife algorithm to obtain those wavelengths. Calibration model quality can be assigned with the high values of determination of coefficient. But in the present research for titratable acidity and fruit firmness, the prediction was not accurate. So this has concluded that for these two parameters the near infrared spectroscopy has not proved as an efficient tool for rapid analysis [30].

Romano et al. (2016) have determined the technique to know the moisture content present in the mango fruit. During the proximate analysis the effect of high temperature had damaged the chemical bonds and chemical constituents had changed. To protect the sample from thermal damage the present research had applied the near infrared spectroscopy and obtained two different wavelengths that is 532 nm and 738 nm for the moisture determination in a non-invasive way. In this way the present research has reduced the cost, time, man power and industrial problem of the analysis and has increased the accuracy of the method [31].

Omar and Yahaya (2017) have discussed about the different chemmometric approaches to determine the physical properties of popular Malaysian sala mangoes. The physical characteristics have been classified with the weight of the intact mangoes and with the colour indices of the mango. The study was done with reflectance and interactance way of Vis-NIR spectroscopy with multiple linear regression (MLR). Different preprocessing

(MSC and SG) have been performed in terms of getting best results on the original spectral data. For the prediction of weight and colour index of “sala” mango MLR + SG and MLR + MSC were appropriate respectively. The comparison was done with both the reflectance and interactance mode and was concluded that there is a similarity in results. Thus, two methods along with MLR can be taken into the consideration for estimation of physical attributes of mango [32].

Kipruto et al. (2017) have been discussed about a newly invented technology through their research that is NIR and evaporative cooling method. The method was developed to maintain the quality and shelf life of the mangoes with a maintained low temperature and the advantage of this method was to eliminate the use of conventional electrical refrigerator. Two different methods NIR reflectance cooling and evaporative cooling jointly have been applied and different temperatures during the storage were maintained. For NIR storage the temperature was varied from 15.4 ˚C to 18.6 ˚C and for non NIR storage and ambient storage the temperature were varied from 16.4 ˚C to 22.9 ˚C and from 23.6 ˚C to 31.9 ˚C. Also the samples were kept at room condition; in that case the temperature was maintained from 21.6 ˚C to 25.5 ˚C. Total four parameters namely shelf life, fruit firmness, skin color and intact fruit weight has been analyzed during a storage time period with different storage condition. Every case has been reported a huge change in these properties during the storage in non-NIR, room condition and ambient storage. But the results of concerned properties have been reported very less changes as compare to others in case of NIR storage. Thus, the present research has proved the efficiency of the NIR storage condition to maintain the quality of mangoes [33].

Taira et al. (2017) have discussed about the prediction of different biochemical parameters with the non-conventional method. Irwin mango samples had been taken for analysis. Mango is a tropical fruit, which has very high nutritious content. Some of them can be predict extrinsically and some of them intrinsically as well. Both the case, near Infrared spectroscopy is a common tool. With the principle of infrared

spectroscopy a spectrometer has been developed [34].

Anderson et al. (2017) have obtained a regression statistical relation successfully with the application of near infrared spectroscopy. Tw different treatment methods were discussed to improve the dry matter content in the mango fruit. The dry matter content with different stages of harvest was obtained rapidly with NIR spectrometry. The value of coefficient of determination has defined 0.82 in the present study which has proved the efficiency of the rapid determination [35].

Mithun et al. (2017) have applied the visible spectrum 350-1050 nm to discriminate two types of mangoes. One was naturally ripened mangoes and another one was artificially ripened mangoes with some stimulator. The wavelength selection method was done and the range from 390 nm to 700 nm had been identified as the best range to detect the same. A regression model was done to reduce the 216 spectral data points to 36 spectral data points positively. A quality and rapid method was derived from this research to resolve a postharvest industrial problem [36].

Li et al. (2018) have discussed the change in flesh colour for the deterioration in mango fruit. The colour of fruit was identified with the visible-near infrared spectroscopy (350-1000 nm). Two specific cases were taken for the experimentation. For first case the deterioration level was light, but flesh colour of mango had been turned into black. But, in second case the deterioration was obtained so high and was predicted the colour lighter than the original flesh colour of mango fruit. The particular wavelengths were selected from all the range and it was determined that 765 nm and 815 nm were the two important peaks through which the detection was done rapidly [37].

Joomwong et al. (2018) have discussed about the physical and chemical analysis of the mango fruit during three harvest period. The mangoes depending upon these three

harvesting stages were taken as mature, immature and over mature. All these mango samples were taken for the analysis and a portable near infrared spectrometer was used to estimate those properties. During the experimentation total soluble solid (TSS) and titratable acidity (TA) were determined and also a ratio of these two parameter was measured. This study was proved the usefulness of the portable NIR spectrophotometer [38].

Polinar et al. (2019) have been reported about the assessment of near infrared spectroscopy for prediction of total soluble solid and dry matter content of Philippian variety Carabao mangoes. The study was carried out with partial least square regression method and the results were observed at the different period after flowering. The observation was taken at 110^{th}, 115^{th}, 120^{th} and 125^{th} day after flowering. The samples were taken at the hard green stage and the estimation model was produced a nondestructive and robust method to grade and sort the sample very easily [39].

Lakade et al. (2019) have discussed the non-destructive prediction method in their research work to predict the harmful effect of this parameter. Mango ripening is a natural process, but sometimes farmers and wholesalers use the calcium carbide which is an acytelene-releazing chemical. This is repining stimulator, with which mango cannot ripe naturally. As this factor calcium carbide is very harmful for health and arsenic content found in the mango after using it, so use of calcium carbide by the sellers is banned now. But people are using it. Samples were taken from the local market of Pondicherry, India and differentiate the samples with two groups. One group kept as for natural ripening process while another kept for stimulator (i.e. Calcium carbide). Spectral data had been taken from the spectrometer for the analysis. The regression model was also used to terminate the relation between the parameters of naturally and artificially ripe mango [40].

Munawar et al. (2019) have discussed about nondestructive analysis of mango to

estimate Titratable acidity, Vitamin C content and soluble solid content. Four different types of mangoes (i.e. Kent, Cengkir, Kweni and Plamer) had been taken for analysis. Regression model has been developed [41].

Wokadala et al. (2019) has been studied about the nondestructive moisture content estimation. During the solar dying to study the moisture content in specifically four types of mangoes has been taken. For commercial acceptance to civilization moisture content of a fruit is an essential parameter to maintain. This parameter keeps the aroma and textural properties of the fruit properly. Less moisture helps to keep the long shelf-life of a fruit. They have performed the chemical (destructive) method as well as the chemmometric approach to predict the parameters. A model was developed using the regression analysis. The evaluated parameters are TSS and TA [42].

Cheng et al. (2019) have discussed about three different mango drying methods and the final result of dehydration of mango slices were analyzed with the near infrared spectroscopy. The present research also has been obtained the application of ANOVA simultaneous component analysis (ASCA) on the near infrared spectral data to determine the same. Traditional sun drying method, tunnel drying method and electrical oven drying method had been used in this research. Among these three every method had come out with some learning outcomes. The sun drying method was responsible for some change in composition in the mango pulp. Chlorophyll and carotenoid are the important pigments present in the mango and these pigment concentrations had not affected during the oven drying method. But among these three drying approaches the tunnel drying method was proved as more efficient with better results [43].

Anderson et al. (2019) have determined rapid and non-destructive assessment of fruit dry matter content. The study has been used the near infrared spectroscopy along with partial least square regression analysis for estimation of dry matter content. It has

concluded that the dry matter content had been increased with the maturity of the fruit and the harvesting period of the fruit. Dry matter content is linearly correlated with these two parameters [44].

Munawar et al. (2019) have estimated vitamin C content in mango fruit. Mango is a good source of vitamin C and it was measured with 2.6 Dichlorophenolindophenol solutions with a destructive approach. But this measurement can be done with a rapid and robust method with NIR spectroscopy (1100-2500 nm). This method with two statistical analysis process that is Principal Component Analysis (PCA) and partial least square analysis (PLS) for determination of vitamin C. Both the cases had given a good result and had proved as an efficient tool (as determination coefficient is more than 0.9 for both the case) along with NIR spectroscopy for determination of vitamin C [45].

Gabriels et al. (2019) have been estimated the fruit firmness and flesh colour of the mango fruit to detect the fruit quality index. These two parameters were indicated the fruit quality and estimation of these two parameters of some Brazilian mangoes was done in this present study. The study has been depicted the role of near infrared spectroscopy for these internal quality analysis of mango [46].

Anderson et al. (2019) have developed a regression model using partial least square method and NIR spectroscopy. This study has concluded with a regression model for dry matter content of mango with the regression coefficient 0.82 and error 0.52. This was good estimation model and dry matter content was predicted rapidly with this. The rapid method was applied in the present research for determination of dry matter content in different treatments of mango orchards. Impact of water denial method during harvesting had been discussed and also had produced that it was increased the dry matter of mango [47].

Saksangium and Sirisomboon (2020) have discussed about the usefulness of the FT-NIR instrument by the precision of two different parameters that is pH and TSS of mango. Precision of these two parameters has evaluated with the reproducibility and repeatability values. The values of repeatability and reproducibility of FT-NIR and UV-Vis spectrometer are 0.00323, 0.03561 and 0.00191, 0.00529 respectively. With these precision indicators a prediction model was developed with the value of determination of coefficient for TSS and pH as 0.9825 and 0.9504 respectively [48].

A robust and quick nondestructive method was developed by **Shah et al. (2020)** and the determination of maturity of mango fruit has done with the study. First the authors have calculated the results with pre-calibrated Kensington Pride (KP) model for the maturity determination of mango. This model was pre-calibrated with some local varities, but it was not applicable for all the varities. So, a calibration of regression model was performed using chemometric studies. Chemometric was performed with the reference data and PLS regression. A portable NIR instrument for field based experiment was developed using least square support vector machine learning model (LSSVM) [49].

Hayati et al. (2020) have presented work to describe an improved near-infrared spectral dataset used to increase the accuracy of near-infrared models in predicting mango fruit quality metrics. Total acidity (TA) and vitamin C, which relate to the primary internal characteristics of fruits, are the two quality parameters that have been cited. In the wavelength range of 1000 to 2500 nm, near infrared (NIR) spectra data were collected and stored as absorbance spectral data. Following that, these data were improved using a variety of techniques, including multiplicative scatter correction (MSC), baseline linear correction (BLC), and their combination (MSC+BLC). TA and vitamin C prediction models were created using the most popular method, partial least square regression (PLS), based on raw and enhanced spectral data, respectively. Performances of predictions can be measured with the discussed study [50].

A significant tropical fruit called the mango (Mangifera indica L.) has the potential to experience interior physiological issues when it ripens later. These include jelly seeds, which have a transparent, jelly-like tissue surrounding them that gradually turns into a brown ring encircling them, and black flesh, which has a diffuse brown discoloration covering the seeds. Due to the lack of knowledge regarding the underlying mechanisms and effective management strategies, these illnesses have the potential to cause significant postharvest losses. Determination of the feasibility of portable visual and near-infrared (Vis-NIR) spectrometer to detect mangoes with internal abnormalities such jelly seeds and black flesh after storage has been done by **Mogollon et al. (2020)**. 141 'Keitt' mangoes total, harvested from two At harvest and 30 days later at 12 °C, 141 'Keitt' mangoes from two commercial harvests were measured spectrally on two opposing cheeks. Utilizing logistic, supporting vector machine, functional data, and random forest modeling techniques, classification models were created utilizing spectra data as well as the incidence of jelly seed and black meat after storage. The findings demonstrate that fruit with internal physiological problems, such as jelly seed and black meat, can be predicted at harvest and detected during storage using wavelengths between 550 and 650 nm. Internal problems could not be separated from one another, nevertheless. According to the spectral data, healthy fruit reflect light more intensely than jelly seeds and dark-colored fruit. The Logistic and LDA model development approaches produced the best classification models [51].

Due to the wide range in fruit maturity stage at arrival, this step is challenging. There were currently no effective methods for quickly and nondestructively observing and managing the mango ripening process. Fruit firmness is a key sign of mango ripeness. The firmness of each mango that was undergoing ripening was predicted by **Mishra et al. (2020)** using a portable visible near-infrared (VNIR) spectrometer (400-1130 nm). The 'Kent' mango's ripening was observed for 10 days at a temperature of 20 °C and a

relative humidity (RH) of 85%. On two different sides of the fruit, the NIR spectrum and fruit firmness (as assessed by the AWETA acoustic firmness analyzer) were measured every other day. For the purpose of determining the significant wavelengths responsible for predicting mango firmness, interval partial-least square (iPLSR) regression was performed. The VNIR spectra changed as a result of the mangoes' changing firmness, according to the results. Compared to PLSR without pre-selecting wavelengths, the model based on chosen wavelengths performed noticeably better. According to iPLSR-based regression, calibration and prediction were correlated with R2c = 0.75 and R2p = 0.75, respectively, and their respective root means squared errors were 6.02 Hz2g2/3 and 5.92 Hz2g2/3. Over 12% more R2p and a 14% reduction in prediction error were achieved by the iPLSR model compared to the conventional PLSR model. The model's predictions showed how the hardness changed over the course of the entire ripening trial. When mangoes are ripening, non-destructive access to their firmness can help optimize the process so that it better meets consumer demand [52].

The failure of models occurs frequently in near-infrared (NIR) spectroscopy of fresh fruit because of external impacts brought on by variations in the physical, chemical, and climatic variables. Correction techniques are needed, which involve measuring reference samples under all possible scenarios before re-modelling is done. However, it is frequently challenging to measure standard samples in the real world. To address this, dynamic orthogonalization projection (DOP) and domain adaption (DA), two distinct ways to adjust for environmental impacts without standard sample measurements, are discussed by **Mishra et al. (2020)** and, for the first time, applied to NIR spectroscopy of fresh fruit. For the demonstration, four separate case studies that were picked for their significance and recurrence in the NIR spectroscopy domain were employed. The first scenario included changing a model such that it could continue to forecast accurately when applied to spectra from a second, related sensor. The second instance involved the rectification of temperature effects brought on by

sensor heating. The third and fourth examples focused on keeping the performance of the model for multi-season fruit quality prediction models for mangos and apples. The goal in each situation was to overcome the difficulties without using fresh criteria for measuring. The findings demonstrated that both DOP and DA enhanced model performances in every situation. There was a 98% reduction in prediction bias and a 66% reduction in root mean squared error (RMSE) of prediction, as well as an increase in R^2_p of up to 31%. The fundamental advantage of DOP and DA approaches in NIR spectroscopy is the reduced need for standard measurements, which makes the models accessible, transferable, and reproducible [53].

Goisser et al. (2021) have discussed about portable near-infrared (NIR) spectrometers—also known as food scanners—perform in terms of their ability to accurately forecast key quality indicators for fruits and vegetables. Therefore, destructive measurements of the related quality parameters (sugar content, dry matter, relative water content) were made in conjunction with food scanner measurements on a variety of fruits from the fruit and vegetable range. The dry matter content of apples, avocados, blueberries, table grapes, and tangerines were assessed in this study, and the results showed cross-validation (R^2) correlations of up to 0.95, 0.87, 0.94, 0.92, and 0.92, respectively. Furthermore, cross validation correlations (R^2) of up to 0.95, 0.84, 0.80, 0.75, 0.95, 0.93, and 0.87 were found when food scanner spectra were used to predict the sugar content of blueberries, kiwi, mango, persimmon, table grapes, tangerines, and tomatoes. Additionally, a correlation of R^2= 0.91 was found for the relative water content of ginger. The findings demonstrate that these traits may be accurately predicted non-destructively using three commercial food scanners: SCiOTM, F-750 Produce Quality Meter, and H-100F. Thus, to quickly assess fruit quality, food scanners can be utilized as objective measuring devices along the value chain for fruits and vegetables. Additionally, employing arriving goods inspections in the wholesale fruit and vegetable trade, these measuring tools' potential for non-destructive quality assessment is demonstrated [54].

The sweetness of the fruits is one indication of the mango's maturity, which was predicted with polynomial regression by **Ulya et al. (2021)**. A mature avomango has a high level of sweetness that is attributed to its high percentage of total soluble solids (TSS). Near Infra-Red (NIR) spectroscopy is currently used in numerous non-destructive studies to determine the TSS concentration. The level of sweetness of avomangos can be predicted using the spectra data produced by NIR spectroscopy. This study compares a multi-predictor local polynomial regression strategy to multiple polynomial regression in order to predict the level of avomango sweetness. In this study, we split up 120 Avomango samples into two groups: 100 were used as training data and 20 were used as testing data. The multi-predictor local polynomial regression performs better with a mean absolute percentage error (MAPE) for predicting the sweetness level of avomangos of 8.554%, which is classed as a highly accurate prediction [55].

Mishra and Passos (2021) have developed a novel method for utilising chemometrics expertise to enhance deep learning (DL) models for analysing spectral data. To enhance the prediction accuracy of DL models built on spectral data, the technique suggests pre-filtering the outliers using Hotelling's T2 and Q statistics produced with partial least-square (PLS) analysis and spectral data augmentation in the variable domain. The same data is pre-processed using a variety of pre-processing techniques, including standard normal variate, first and second derivatives, and merge of both, to create augmented data. On an actual near-infrared (NIR) data set pertaining to the DM estimation in mango, the approach's effectiveness is proved. The total number of spectra and reference DM measurements in the data set was 11,961. The outcomes demonstrated that eliminating the outliers and enhancing spectral data increased the DL models' capacity for prediction. Furthermore, in terms of good RMSEP values of 0.84%, this method results more accurate DL models and also achieved the minimum value of RMSEP on the mango data set, at 0.79%. Furthermore, the RMSEP dropped

to 0.75% once outliers were eliminated from the test set. Many chemometrics methods can supplement DL models [56].

Shah et al (2021) have investigated a novel non-destructive hand-held fruit ripeness measurement method. The presented approach uses a classifier trained on maturity labels supplied through common DM criteria for the tested mango types to directly classify the maturity state (mature/immature).The estimation of on-tree mango ripeness is crucial for determining when to pick the fruit. A useful indicator of mango maturity and post-harvest quality is dry matter (DM). Existing NIR-based maturity metres use machine learning regressors to forecast the value of a maturity index (such DM, oBrix, or etc.) and then apply a hard threshold to the expected value to determine the fruit's maturity state. The C++ application software's functions include gathering interactance spectra, reducing noise, dimensionality, and maturity state classification. By using samples of mango fruit from different seasons grown on the tree, the performance of the suggested approach is assessed. Both the proposed direct maturity categorization and the indirect maturity assessment published in the literature are compared. According to the test results, indirect maturity assessment with hard thresholds can obtain a maximum accuracy of 55.9%. The test mangoes' maturity condition (mature/immature) was correctly predicted by direct maturity classification using KNN with an accuracy of 88.2% [57].

Zakaria et. al. (2021) have examined the viability of employing artificial neural networks and near-infrared light to non-destructively screen the Harumanis mango for IFR. A photodiode was utilized to obtain the intensity of the NIR light reflected by the Harumanis mango, and a typical NIR light emitting diode with a wavelength of 1000 nm was utilised as the light source to emit NIR light. Local experts used the acoustic method to manually detect IFR early on by flicking their fingers to feel for any abnormalities inside the fruit. In order to categorize the presence of IFR using a neural network, sample data on the NIR Spectroscopy reflectance results of 120 samples

were employed. With an incidence of Insidious Fruit Rot, the Harumanis mango's mean NIR reflectance RGB values are 0.651, 0.465 and 0.458 respectively, while those without IFR RGB values are 0.211, 0.15, and 0.146 respectively [58].

Maraphum et al. (2022) have selected the best wavelength to use for Nam-Dokmai mango quality testing using near infrared (NIR) spectroscopy. NIR spectroscopy has been used in this work to grade management systems for commercial mango export. A near infrared sensor with a wavelength range of 860–1760 nm was used to capture near infrared spectra. The spectrum wavelengths were chosen using a genetic algorithm (GA) and successive projections algorithms (SPA). Additionally, partial least square (PLS) regression was utilized to create the prediction models using the chosen wavelengths. The second derivative was used to determine the ideal pretreatment. With R^2 of 0.66-0.74, RMSEP of 0.72-0.80 °Brix, and RPD of 1.8-2.0, the model with full wavelengths produced the best result. The r 2, RMSEP, and RPD values obtained from the SPA-PLS were 0.43–0.70, 0.77–1.01°Brix, and 1.4–1.9, respectively. The results of the GA-PLS conducted effectively with R^2, RMSEP, and RPD were, respectively, 0.52-0.72°Brix, 0.74-0.96°Brix, and 1.5-1.9°Brix. The result, the 50-variable GA-PLS model, is appropriate for use in determining the soluble solids content (SSC) of mangoes. This model might be employed for screening. Additionally, it did not differ noticeably from the best model. As a result, the authors hypothesized that the GA-PLS prediction model, which has 50 factors, can be used to evaluate SSC in mangoes [59].

Seehanam et al. (2022) have discussed about the non-invasive method of near infrared spectroscopy to identify some internal diseases of mango. For mango growers and exporters, internal problems are a serious issue. The most prevalent signs of tropical fruit cultivation are internal breakdown (IBD) and black-streaked vascular tissue (BSV), particularly in Thailand, where mango is the country's most important agricultural export. Consumers discover these unappealing flaws while cutting into the

fruit because the abnormalities are not visible to the naked eye. At 120 days after flowering (DAF), a total of 64 'Namdokmai Sithong' mangoes were collected and manually meshes using a scale of 1.5 1.5 cm2 on both cheeks (totaling 1112 useable areas: intact (792), IBD (230), and BSV (90). In order to distinguish between healthy and unhealthy flesh, spectral data were obtained between 4000 and 12,500 cm 1 for each mesh region. Following that, classification models utilizing linear discriminant analysis (LDA) and an artificial neural network were created (ANN). The NIR absorbance of IBD meat was lower than that of BSV and undamaged flesh. The LDA model had 86.25% prediction accuracy when separating healthy flesh from unhealthy flesh. It was unable to distinguish between entire meat and IBD, though. The ANN achieved a classification accuracy of 91.37% in the case of non-linear analysis, and the misclassified matrix demonstrated that intact, IBD, and BSV flesh were clearly distinguishable from one another, particularly for BSV flesh. In conclusion, mango internal abnormalities could be discovered by NIRS [60].

Zeb et al. (2022) have used short-wave NIRS (near-infrared spectroscopy) for examination of fruit classification issue. The study focuses on O-H and C-H overtone characteristics of fruit and their correlation with NIRS, which adds a new level of complexity to NIRS-based fruit classification issues. Eleven fruits were employed in this study to examine physical traits including seed thickness, pulp and peel thickness size, including hass, kiwi, cherry, grapes, apples, mango, orange, melon, plum, loquat and apricot. For all of the aforementioned fruit, NIR spectral data is gathered using the industry-recognized F-750 fruit quality meter (wavelength range 300-1100nm). Utilizing spectral feature vectors, various shallow machine learning architectures were trained to categorize fruits. Using four different features with wavelengths of 770nm, 840nm, 910nm, and 960nm after first employing 83 features vectors with a range of 725-975nm (3nm-resolution) (corresponding to O-H and C-H overtone features). For the input set of 83 spectral features and four featers, the QDA classifier achieved a cross-validation accuracy and a test data accuracy of 100% and 93.02% and 97.1%

and 90.38% respectively. The findings show that short-wave NIR radiation absorptivity, particularly with regard to O-H and C-H overtone characteristics, is a key factor in fruit classification. The primary wavelength ranges of an LED-based device are 770, 840, 910, and 960 nm. LEDs can be utilized in automation-related applications [61].

There are several appropriate classical machine learning techniques discussed, including classifiers, feature extraction techniques, and feature reduction techniques. A suggestion of machine learning algorithms addresses the issues of categorizing mangoes into mature, non-mature, and matured classes further into ripened and under-ripened. However, a hierarchical multi-classifiers fusion strategy is created to classify ripe mangoes into four categories: perfectly ripe, overripe, overripe with black spots on the skin, and without black spots on the skin. Additionally, a technique for choosing the optimal NIR spectroscopic wavelengths in order to propose a non-destructive method of identifying internal mango flaws was also introduced by **Guru et al. (2022)**. It is also attempted to categorize ripened mangoes into naturally ripened and artificially ripened mangoes, which is the most challenging problem, and an appropriate classifier is provided. The deep learning alternatives to each problem being tackled using traditional machine learning techniques are also mentioned. On moderately significant datasets of mango photos produced during the course of our research, the results of rigorous experimentation carried out to show the effectiveness of our methodologies are presented. There is a full explanation of datasets as well as the challenges encountered in creating them. A comparison of several methods, including those based on deep learning, is offered. Future research possibilities in related fields are also examined. Overall, the goal of this work is to present an integrated approach to handling mangoes after harvest, with a focus on automating the grading and sorting procedures [62].

References

1. S. Saranwong, J. Sornsrivichai, S. Kawano, *J. Near Infrared Spectrosc*., **9**, 287-295 (2001).
2. S. Saranwong, J. Sornsrivichai, S. Kawano, *J. Near Infrared Spectrosc.*, **11**, 283-293 (2003).
3. S. Saranwong, J. Sornsrivichai, S. Kawano, *J. Near Infrared Spectrosc*., **11**, 175-181 (2003).
4. S. Saranwong, J. Sornsrivichai, S. Kawano, *Postharvest Biol. Technol*., **31**, 137–145 (2004).
5. B. Mahayothee, W. Mühlbauer, S. Neidhart, M. Leitenberger and R. Carle, *Acta Hortic*., **645**, 581-588, 2004.
6. K. B. Walsh, M. Golic, C. V. Greensill, *J. Near Infrared Spectrosc.*, **12**, 141-148 (2004).
7. J. Ahmed, H. S. Ramaswamy and N. Hiremath, *Int. J. Food Science Technol.*, **40**, 885-895 (2005).
8. A. Gundurao,H. S. Ramaswamy and J. Ahmed, *Int. J. Food Prop.*, **14**, 1018-1036 (2011).
9. S. R. Delwiche,, W. Mekwatanakarn and C. Y. Wang, *Horttechnology*., 18, 410-416 (2008).
10. M. Valente, R. Leardi, G. Self, G. Luciano, P. J. Pierre, *J. Food Eng*., **94**, 7–13 (2009).
11. J. Yu and Y. He, Int. *Federation Inf. Process.*, **293**, 89-95 (2009).
12. P. P. Subedi, K. B. Walsh, *Postharvest Biol. Technol*., **51**, 297-304 (2009).
13. M. Nagle, B. Mahayothee, P. Rungpichayapichet, S. Janjai, J. Müller, *Sci. Hortic*., **125**, 771-774 (2010).
14. S. Saranwong, W. Thanapase, N. Suttiwijitpukdee, R. Rittiron, S. Kasemsumran, S. Kawano, *J. Near Infrared Spectrosc*., **18**, 271-280 (2010).
15. A. Gundurao, H. S. Ramaswamy and J. Ahmed, *Int. J. Food Prop.*, **14**, 1018-1036 (2011).

16. S. Saranwong, R. P. Haff, W. Thanapase, A. Janhiran, S. Kasemsumran, S. Kawano, *J. Near Infrared Spectrosc*., **19**, 55-60, (2001).
17. S. N. Jha, P. Jaiswal, K, Narsaiah, M. Gupta, R. Bhardwaj, A. K. Singh, *Sci. Hortic*., **138**, 171–175 (2012).
18. U. Kitthawee, S. Pathaveerat, D. C. Slaughter, E. J. Mitcham, In *Asia Pacific Symposium on Postharvest Research, Education and Extension* **943**, 287-296, (2010).
19. A. I. Basem, M. I. Fadhel, T. Al-kharazi, M. Ali and S. Miloud, *J. Food Sci. Technol., **51***, 3244-3252 (2014).
20. A. B. Izneid and T. Al-kharazi, *Proceeding of the 2013 IEEE International Conference on RFID Technologies and Applications, 4 – 5 September, Johor Bahru, Malaysia.*
21. R. P. Haff, S. Saranwong, W. Thanapase, A. Janhiran, S. Kasemsumran and S. Kawano, *Postharvest Biol. Technol., **86***, 23-28 (2013).
22. C. Watanawan, T. Wasusri, V. Srilaong, C. Wongs-Aree and S. Kanlayanarat, *Int. Food Res. J.*, **21**, 1073 (2014).
23. N. Vélez Rivera, J. Gómez-Sanchis, J. Chanona-Pérez, J. J. Carrasco, M. Millán-Giraldo, D. Lorente, S. Cubero, *J. Blasco, Biosys*. Eng., **122**, 91-98 (2014).
24. P. Theanjumpol, K. Wongzeewasakun, N. Muenmanee, S. Wongsaipun, C. Krongchai, V. Changrue, D. Boonyakiat and S. Kittiwachana, *Postharvest Biol. Technol*., **153**, 13-20 (2019).
25. Y. Y. Pu and D. W. Sun, *Food Chem*., **188**, 271-278 (2015).
26. A. A. Munawar, D. V. Horsten, J. K. Wegener, E. Pawelzik and D. Morlein, *Eng. Agric. Environ. Food*, **xxx**, 1-8 (2016).
27. V. E. Nambi, K. Thangavel, K. A. Rajeswari, A. Manickavasagan, and V. Geetha, *Postharvest Biol. Technol.*, **117**, 152-160 (2016).
28. Y. Y. Pu and D. W. Sun, Innov. *Food Sci. Emerg. Technol*., **33**, 348-356 (2016).
29. P. Rungpichayapichet, B. Mahayothee, M. Nagle, P. Khuwijitjaru and J. Müller, *Postharvest Biol. Technol*., **111**, 31-40 (2016).

30. E.J.N. Marques, S. T. De Freitas, M. F. Pimentel and C. Pasquini, *Food Chem.*, **197**, 1207-1214, (2016).
31. G. Romano, M. Nagle, J. Müller, *Innov. Food Sci. Emerg. Technol*., **33**, 498-505, (2016).
32. A. F. Omar and O. K. M. Yahaya, "Spectrometry Evaluation of B10 Averrhoa carambola L. and Sala Mango Physical Attributes". In *9th International Conference on Robotic, Vision, Signal Processing and Power Applications* (Springer, Singapore, 2017), pp. 655-662.
33. K. M. Kipruto, *Agric. Eng. Int.: CIGR J.*, **19**, 162-168 (2017).
34. E. Taira, S. Nakamura, R. Hiyan, H. Honda, M. Ueno, *Eng. Appl. Sci. Res*., **44**, 189-192 (2017).
35. N. T. Anderson, P. P. Subedi, K. B. Walsh, *Sci. Hortic*., **226**, 316-321 (2017).
36. B. S. Mithun, M. Mondal, H. Vishwakarma, S. Shinde, S. Kimbahune, in *Proceedings of SPIE - The International Society for Optical Engineering*, 10217, (2017).
37. L. Li, Y. Li, Y. Peng, in *ASABE 2018 Annual International Meeting*, (2018).
38. A. Joomwong, J. Joomwong, P. Noimanee, P. Theanjumpol, *Acta Hortic*., **1213**, 225-228 (2018).
39. Y. Q. Polinar, K. F. Yaptenco, E. K. Peralta and J. U. Agravante, *Agric. Eng. Int.: CIGR J.*, **21**, 209-219 (2019).
40. A. J. Lakade, V. Venkataraman, R. Ramasamy, P. H. Shetty, *Food Addit. Contam. Part A*, **36**, 989-995 (2019).
41. A. A. Munawar, R. Hayati, D. Wahyuni, *INMATEH – Agric. Eng*., **58**, 1-12 (2019).
42. O. C. Wokadala, C. Human, S. Willemse, N. M. Emmambux, *J. Food Meas. Charact*., **14**, 790-198 (2019).
43. W. Cheng, K. M. Sørensen, R. J. Mongi, B. K. Ndabikunze, B. E. Chove, D. W. Sun, S. B. Engelsen, *LWT*, **112**, 108214 (2019)
44. N. T. Anderson, P. P. Subedi, K. B. Walsh, *Acta Hortic*., **1244**, 123-128 (2019).

45. A. A. Munawar, R. Hayati, D. Wahyuni, *INMATEH – Agric. Eng.*, **58**, 1-12 (2019).
46. S. H. E. J. Gabriels, B. Brouwer, H. de Villiers, E. Westra, E. J. Woltering, *Acta Hortic*., **1256**, 289-294 (2019).
47. N. T. Anderson, K. B. Walsh, P. P. Subedi, Acta Hortic., **1244**, 69-74 (2019).
48. N. Saksangium and P. Sirisomboon, "The overall precision test for near infrared scanning and reference method for the determination of soluble solids content and pH of mango for processing factory". In *E3S Web of Conferences* 187, (EDP Sciences, 2020) p. 04003.
49. S. S. A. Shah, A. Zeb, W. S. Qureshi, M. Arslan, A. U. Malik, W. Alasmary and Alanazi, E., *Infrared Phys. Technol.*, **111**, 103479 (2020).
50. R. Hayati, A. A. Munawar and F. Fachruddin, *Data in brief*, **30**, 105571 (2020).
51. R. Mogollón, C. Contreras, M. L. da Silva Neta, E. J. N. Marques, J. P. Zoffoli and S. T. de Freitas, *Postharvest Biol. Technol.*, **167**, 111251 (2020).
52. P. Mishra, E. Woltering and N. El Harchioui, *Infrared Phys. Technol.*, **110**, 103459 (2020).
53. P. Mishra, J. M. Roger, D. N. Rutledge and E. Woltering, *Postharvest Biol. Technol.*, **170**, 111326 (2020).
54. S. Goisser, S. Wittmann and H. Mempel, *Landtechnik*, **76**, 52-67 (2021).
55. M. Ulya, N. Chamidah and T. Saifudin, "Predicting the sweetness level of avomango (Gadung Klonal 21) using multi-predictor local polynomial regression". In *IOP Conference Series: Earth and Environmental Science* 733, (IOP Publishing, 2021), p. 012009.
56. P. Mishra and D. Passos, *Chemom. Intell. Lab. Syst.*, **212**, 104287 (2021).
57. S. S. A. Shah and A. Zeb, W. S. Qureshi, A. U. Malik, M., Tiwana, K. Walsh, A. Muhammad, A. Waleed, and E. Alanazi, *Infrared Phys. Technol.*, **115**, 103639 (2021).
58. N. S. Zakaria, A. H. Abdullah, F. A. Saad, M. A. Hassan, M. A. Bakar, A. M. Saad, S. Sudin M. F. Ibrahim, "Potential of Near-Infrared (NIR) spectroscopy

technique for early detection of Insidious Fruit Rot (IFR) disease in Harumanis mango". In *Journal of Physics: Conference Series* 2107, (IOP Publishing, 2021) p. 012014.

59. K. Maraphum, A. Ounkaew, P. Kasemsiri, S. Hiziroglu and J. Posom, *Eng. Appl. Sci. Res.*, **49**, 119-126 (2022).
60. P. Seehanam, P. Chaiya, P. Theanjumpol, C. Tiyayon, O. Ruangwong, T. Pankasemsuk, K. Nakano, S. Ohashi and P. Maniwara, *Hortic. Environ. Biotechnol.*, 1-11 (2022).
61. 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)* (IEEE, 2022) pp. 193-196.
62. D. S. Guru, A. Raghavendra and M. K. Rao, "Post-harvest Handling of Mangoes: An Integrated Solution Using Machine Learning Approach", in *International Conference on Computer Vision and Image Processing*, (Springer, Cham, 2022) pp.-243-253.

Chapter-3: Objective and Methodology

Chapter-3

Objective and Methodology

This chapter focuses on the research gap which is obtained from all the present literature and it also covers the objectives of the present work along with the methodology to achieve that.

3.1. Research Gap

After surveying the literature about all the destructive as well as non-destructive analysis to predict the different extrinsic (i.e. size, weight, width, flesh color etc.) and intrinsic (i.e. firmness, total soluble solids, titratable acidity, Vitamin content etc.) parameters of mango, it is found that very few researches had been done for analyze the different nutritious quality parameters of mango. “King of fruits” is having richest nutritious value with high dietary fiber content, good vitamin content and good source of water and energy (Kcal). Maximum researchers have followed a trend to study the Total Soluble Sugar (TSS), reducing and non-reducing sugar, Soluble Sugar Content (SSC), pH, ^{o}brix value, titratable acidity of citric acid and moisture content of mangoes through NIR spectroscopy. Many researches had been reported for analyzing the quality parameter by conventional method but it is harmful to the nature, many chemicals have used to do the property analysis. So, it is important to do this same through a green and non-destructive way. There are more parameters like protein content, carbohydrate content, crude fiber content, ash content and viscoelastic properties can be studied. Also some volatile compounds which are responsible for the aroma of the mango can be determined.

Very few research works has been reported on destructive as well as non-destructive analysis for checking its quality parameters in terms of different physical, chemical and rheological properties of Indian mangoes. Thus, a research can be taking out to determine the different physical, chemical and rheological properties and for establishment of a regression model, to determine the different chemical properties in a time and cost

effective manner, using near infrared spectroscopy and chemmometric approach for quality prediction of Indian mangoes.

3.2.Research Objective

1. To investigate the physical, chemical, rheological characteristics of mango pulp and peel using standard conventional methods of Association of Official Agricultural Chemists (AOAC).

2. Classification of mangoes on basis of region by Principle Component Analysis (PCA) related to above mentioned properties using nondestructive techniques.

3. Development of a cost and time effective regression model based on Partial Least square (PLS) regression and Principle Component Regression (PCR) for determination of different parameters of mango using Infrared and chemical analysis.

3.3. Research Methodology

The experimental material has been considered as Indian mango cultivars and fifty one fully matured ripe mangoes have been collected from the local mango farm and local markets (Table 3.1 and figure 3.2) of five regions of India that is Bihar; Maharashtra; Punjab; Uttar Pradesh and West Bengal. The experimentation was carried out at the Food and nutrition laboratory, School of biosciences and bioengineering, Lovely Professional University, Punjab, Research Laboratory, School of Agriculture and Food Innovative Research Laboratory, Advanced Analytical Laboratory, Kolkata. The physical and biochemical and rheological attributes of the fruit have been recorded in the triplicate for more accurate results and take the mean finally.

3.3.1. Physical Parameters

Intact mango fruit was taken for the physical parameter analysis. Three basic parameters i.e. major diameter, intermediate diameter and minor diameter has been calculated with the Vernier calipers. Other physical parameters like geometric mean diameter, arithmetic mean diameter, surface area and sphericity has been determined with some mathematical formulations with the help of three basic diameter values.

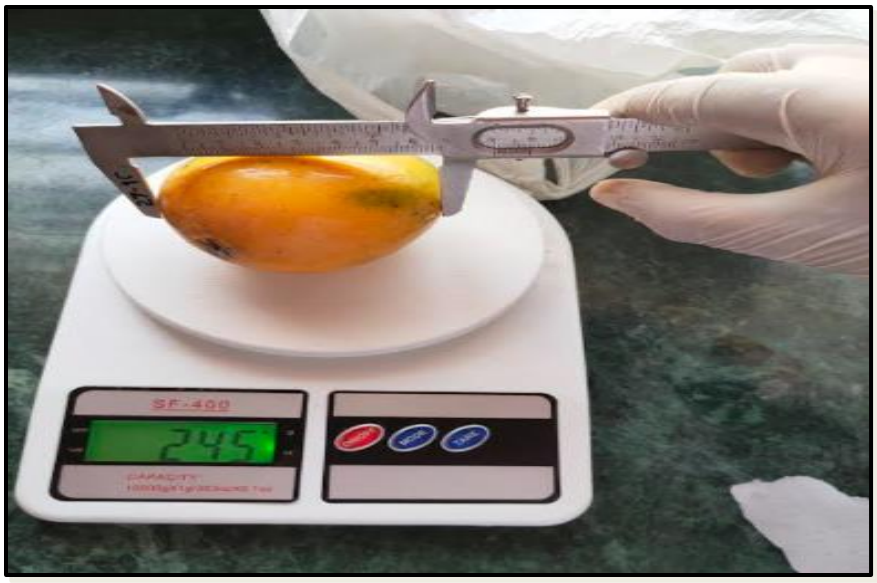


Figure 3. 1 Physical measurement of intact mango fruit using vernier calipers

3.3.2. Sample Preparation

After completing the physical characteristic determination the fruits have thoroughly washed with luke warm water to remove the gummy sap from the mango and cleaned with a muslin cloth. Mango skin of the mango fruit was removed by a hand peeling knife. Pulp has been extracted and peel and kernel have been removed. The average weight of peel, kernel and pulp was recorded before making the homogenized sample. The homogenized solution of mango pulp has been made with the Lifelong Power pro LLMG02 mixer grinder for future analysis. The speed switch was set to number II and the pulp blended for 5 minutes. The homogenized sample

Table 3. 1 State wise and verity wise collected Indian mango samples (Total 51 samples) and their physical properties

Sample	State	Variety	Sample ID	Intact weight (gm)	seed weight (gm)	Peel weight (gm)	Pulp Weight (gm)	Major axis (a) (cm)	Minor axis (b) (cm)	Intermediet axis (C) (cm)	Size (cm)	Surface area (cm^2)	Sphericity (%)
S-2	Punjab	Dusheri	PD	177.000	27.000	42	108	7.77	5.65	6.64	6.630	138.089	0.853
S-3	Punjab	Dusheri	PD	200.000	30.000	51	119	7.53	5.84	6.91	6.723	141.968	0.893
S-4	Maharashtra	Alphanso	MA	229.000	49.000	50	130	9.57	5.64	6.74	7.139	160.067	0.746
S-5	Maharashtra	Imampasand	MI	194.000	42.000	31	121	7.58	6.25	6.38	6.711	141.465	0.885
S-6	West Bengal	Golapkhas	WG	338.000	58.000	102	178	11.06	6.54	7.73	8.238	213.178	0.745
S-7	Maharashtra	Imampasand	MI	179.000	39.000	32	108	7.73	5.88	6.07	6.510	133.116	0.842
S-8	Maharashtra	Alphanso	MA	286.000	54.000	64	168	9.36	6.26	7.43	7.579	180.424	0.810
S-9	West Bengal	Golapkhas	WG	301.000	48.000	89	164	11.55	6.33	7.03	8.010	201.540	0.694
S-10	Maharashtra	Alphanso	MA	259.000	47.000	63	149	10.07	5.77	7.67	7.638	183.260	0.759
S-11	Maharashtra	Imampasand	MI	172.000	38.000	43	91	7.73	5.64	5.97	6.385	128.043	0.826
S-12	Maharashtra	Imampasand	MI	192.000	39.000	38	115	8.33	5.64	8.27	7.297	167.245	0.876
S-13	West Bengal	Golapkhas	WG	212.000	37.000	51	124	9.57	5.23	6.84	6.996	153.714	0.731
S-14	West Bengal	Golapkhas	WG	297.000	53	81	163	10.63	6.42	7.57	8.024	202.229	0.755
S-15	Maharashtra	Imampasand	MI	246.000	45	60	141	9.85	5.88	7.57	7.597	181.276	0.771
S-17	West Bengal	Golapkhas	WG	231.000	46	61	124	10.05	5.64	6.83	7.288	166.844	0.725

Sample	State	Variety	Sample ID	Intact weight (gm)	seed weight (gm)	Peel weight (gm)	Pulp Weight (gm)	Major axis (a) (cm)	Minor axis (b) (cm)	Intermediet axis (C) (cm)	Size (cm)	Surface area (cm^2)	Sphericity (%)
S-18	Maharashtra	Bombay Green	MBG	228.000	41	83	104	9.25	6.36	7.26	7.531	178.139	0.814
S-19	Maharashtra	Bombay Green	MBG	250.000	46	63	141	9.55	5.88	7.17	7.384	171.265	0.773
S-20	West Bengal	Surmaiapally	WSU	354.000	59	29	266	11.35	6.93	7.85	8.515	227.755	0.750
S-21	West Bengal	Surmaiapally	WSU	389.000	69	32	288	12.42	7.27	7.48	8.774	241.791	0.706
S-22	West Bengal	Surmaiapally	WSU	295.000	47	30	218	11.67	6.34	6.76	7.938	197.912	0.680
S-23	West Bengal	Langra	WL	308.000	43	49	216	9.85	6.83	6.38	7.543	178.725	0.766
S-24	West Bengal	Langra	WL	295.000	36	39	220	9.32	7.05	7.29	7.824	192.290	0.840
S-25	West Bengal	Langra	WL	307.000	38	39	230	9.52	6.92	7.32	7.842	193.154	0.824
S-26	West Bengal	Langra	WL	277.000	36	43	198	9.36	6.66	6.93	7.560	179.497	0.808
S-27	West Bengal	Kalapahar	WK	260.000	37	53	170	10.22	6.23	6.73	7.539	178.527	0.738
S-28	West Bengal	Kalapahar	WK	235.000	46	48	141	11.55	5.74	6.35	7.495	176.433	0.649
S-29	West Bengal	Amrapali	WAM	288.000	43	59	186	11.32	6.46	6.64	7.860	194.044	0.694
S-30	West Bengal	Amrapali	WAM	235.000	45	50	140	11.76	5.92	6.18	7.549	179.011	0.642
S-31	West Bengal	Himsagar	WH	341.000	52	71	218	9.79	7.56	7.97	8.387	220.924	0.857
S-32	West Bengal	Himsagar	WH	230.000	42	44	144	8.54	6.65	6.27	7.088	157.797	0.830

Sample	State	Variety	Sample ID	Intact weight (gm)	seed weight (gm)	Peel weight (gm)	Pulp Weight (gm)	Major axis (a) (cm)	Minor axis (b) (cm)	Intermediet axis (C) (cm)	Size (cm)	Surface area (cm^2)	Sphericity (%)
S-33	West Bengal	Himsagar	WH	198	31	48	119	9.55	6.32	5.58	6.957	152.045	0.729
S-34	West Bengal	Himsagar	WH	203	39	41	123	9.63	6.57	5.63	7.089	157.834	0.736
S-35	UttarPradesh	Dusheri	UD	204	26	49	129	10.73	5.36	6.23	7.103	158.454	0.662
S-36	UttarPradesh	Dusheri	UD	245	43	57	145	11.27	5.97	6.32	7.520	177.614	0.667
S-37	UttarPradesh	Dusheri	UD	201	33	51	117	11.56	5.45	5.91	7.194	162.566	0.622
S-38	UttarPradesh	Dusheri	UD	171	29	36	106	10.27	5.26	5.87	6.819	146.060	0.664
S-39	UttarPradesh	Dusheri	UD	199	29	42	128	11.14	5.57	5.85	7.133	159.833	0.640
S-40	Bihar	Himsagar	BH	376	53	62	261	11.16	7.28	8.38	8.797	243.087	0.788
S-41	Bihar	Himsagar	BH	402	55	56	291	10.68	7.22	8.27	8.607	232.708	0.806
S-42	Bihar	Kishenbhog	BKI	616	85	101	430	10.56	8.58	9.68	9.572	287.797	0.906
S-43	Bihar	Kishenbhog	BKI	731	66	114	551	11.72	8.76	10.28	10.181	325.600	0.869
S-44	Bihar	Kishenbhog	BKI	507	54	68	385	10.77	7.97	9.06	9.196	265.625	0.854
S-45	UttarPradesh	Chausa	UC	325	53	82	190	12.43	6.56	7.15	8.354	219.207	0.672
S-46	UttarPradesh	Chausa	UC	328	46	76	206	12.42	6.37	7.24	8.305	216.638	0.669
S-47	Punjab	Safeda	PS	272	63	48	161	10.82	6.12	6.86	7.687	185.611	0.710
S-48	Punjab	Safeda	PS	315	83	56	176	11.64	6.82	6.69	8.098	205.989	0.696
S-49	UttarPradesh	Langra	UL	224	35	44	145	11.48	5.78	6.27	7.465	175.049	0.650

S-50	UttarPradesh	Langra	UL	229	45	50	134	11.91	4.64	6.07	6.948	151.640	0.583
S-51	UttarPradesh	Fazli	UF	290	38	44	208	9.65	6.64	7.18	7.720	187.189	0.800
S-52	UttarPradesh	Fazli	UF	298	41	43	214	9.89	6.64	7.32	7.834	192.746	0.792
S-53	UttarPradesh	Fazli	UF	303	35	40	228	9.34	7.29	7.37	7.947	198.349	0.851

Table 3. 2 State wise and verity wise collected Indian mango samples (Total 51 samples) and their chemical properties

Sample no.	State	Variety	Sample ID	Moisture (%)	Ash (%)	Crude fat (%)	Crude protein (%)	Crude fibre (%)	pH	Carbohydrate (%)	TSS (%)	Energy (kcal/100 g)
S-2	Punjab	Dusheri	PD	89.115	0.335	0.100	4.815	0.510	4.300	5.125	10.000	40.660
S-3	Punjab	Dusheri	PD	88.025	0.205	0.200	5.050	0.590	4.430	5.930	10.000	45.720
S-4	Maharashtra	Alphanso	MA	78.095	2.145	0.200	5.325	0.510	4.045	13.725	10.000	78.000
S-5	Maharashtra	Imampasand	MI	77.190	0.525	0.120	5.405	0.510	3.295	16.250	11.000	87.700
S-6	West Bengal	Golapkhas	WG	87.910	0.535	0.000	5.150	0.545	3.695	5.860	11.000	44.040
S-7	Maharashtra	Imampasand	MI	84.590	0.248	0.275	5.050	0.535	3.385	9.302	11.000	59.883
S-8	Maharashtra	Alphanso	MA	78.860	0.715	0.150	5.350	0.500	3.645	14.425	10.000	80.450
S-9	West Bengal	Golapkhas	WG	87.430	0.575	0.185	6.315	0.500	3.090	4.995	10.000	46.905
S-10	Maharashtra	Alphanso	MA	81.310	0.42	0.590	5.720	0.505	3.210	11.455	12.000	74.010
S-11	Maharashtra	Imampasand	MI	86.800	0.95	0.375	5.355	0.505	3.705	6.015	11.000	48.855
S-12	Maharashtra	Imampasand	MI	80.850	0.455	0.465	6.250	0.560	2.740	11.420	16.000	74.865
S-13	West Bengal	Golapkhas	WG	80.050	0.445	0.335	6.050	0.530	4.465	12.590	10.000	77.575
S-14	West Bengal	Golapkhas	WG	86.870	0.283	0.535	5.150	0.500	4.025	6.662	12.000	52.063
S-15	Maharashtra	Imampasand	MI	80.995	0.525	0.495	5.450	0.470	3.350	12.065	16.000	74.515
S-16	Maharashtra	Imampasand	MI	86.735	2.111	0.455	5.250	0.585	3.250	4.864	12.000	44.551

Sample no.	State	Variety	Sample ID	Moisture (%)	Ash (%)	Crude fat (%)	Crude protein (%)	Crude fibre (%)	pH	Carbohydrate (%)	TSS (%)	Energy (kcal/100 g)
S-17	West Bengal	Golapkhas	WG	87.420	0.375	0.475	4.650	0.530	2.975	6.550	17.000	49.075
S-18	Maharashtra	Bombay Green	MBG	86.500	0.635	0.330	5.150	0.545	3.050	6.840	16.000	50.93
S-19	Maharashtra	Bombay Green	MBG	80.890	0.555	0.400	4.750	0.505	3.715	12.900	17.000	74.200
S-20	West Bengal	Surmaiapally	WSU	82.985	0.425	0.295	6.350	0.605	4.235	9.340	12.000	65.415
S-21	West Bengal	Surmaiapally	WSU	81.140	0.325	0.290	6.050	0.635	4.030	11.560	15.000	73.050
S-22	West Bengal	Surmaiapally	WSU	81.250	0.325	0.275	6.050	0.645	3.985	11.455	10.000	72.495
S-23	West Bengal	Langra	WL	77.585	0.455	0.315	5.450	0.550	4.495	15.645	10.000	87.215
S-24	West Bengal	Langra	WL	82.380	7.85	0.275	6.050	0.635	4.110	2.810	17.000	37.915
S-25	West Bengal	Langra	WL	80.500	0.365	0.245	5.250	0.545	4.025	13.095	10.000	75.585
S-26	West Bengal	Langra	WL	81.585	0.605	0.240	5.150	0.635	3.935	11.785	15.000	69.900
S-27	West Bengal	Kalapahar	WK	78.675	0.545	0.265	5.450	0.645	3.935	14.420	14.000	81.865
S-28	West Bengal	Kalapahar	WK	80.850	0.427	0.245	4.850	0.585	4.790	13.043	15.000	73.777
S-29	West Bengal	Amrapali	WAM	80.355	0.425	0.275	5.150	0.556	3.900	13.239	14.000	76.031
S-30	West Bengal	Amrapali	WAM	82.100	0.3025	0.285	4.650	0.555	4.300	12.107	14.000	69.595
S-31	West Bengal	Himsagar	WH	82.400	0.445	0.305	4.250	0.600	3.810	12.000	15.000	67.745

Sample no.	State	Variety	Sample ID	Moisture (%)	Ash (%)	Crude fat (%)	Crude protein (%)	Crude fibre (%)	pH	Carbohydrate (%)	TSS (%)	Energy (kcal/100 g)
S-32	West Bengal	Himsagar	WH	86.165	0.325	0.365	3.750	0.615	4.195	8.780	12.000	53.405
S-33	West Bengal	Himsagar	WH	79.410	0.3555	0.425	5.150	0.685	4.160	13.974	15.000	80.323
S-34	West Bengal	Himsagar	WH	82.425	0.345	0.400	5.450	0.635	3.950	10.745	14.000	68.380
S-35	UttarPradesh	Dusheri	UD	76.815	0.365	0.605	6.300	0.595	4.350	15.320	12.000	91.925
S-36	UttarPradesh	Dusheri	UD	77.250	0.285	0.545	7.210	0.645	4.255	14.065	15.000	90.005
S-38	UttarPradesh	Dusheri	UD	80.935	0.325	0.605	7.250	0.635	4.185	10.250	12.000	75.445
S-39	UttarPradesh	Dusheri	UD	78.285	0.635	0.365	5.150	0.515	4.200	15.050	16.000	84.085
S-40	Bihar	Himsagar	BH	87.150	0.425	0.355	4.250	0.605	4.200	7.215	15.000	49.055
S-41	Bihar	Himsagar	BH	87.260	0.265	0.315	5.150	0.545	4.240	6.465	11.000	49.295
S-42	Bihar	Kishenbhog	BKI	82.045	0.538	0.445	5.350	0.695	4.210	10.927	15.000	69.113
S-43	Bihar	Kishenbhog	BKI	78.050	0.315	0.525	7.050	0.675	4.250	13.385	15.000	86.465
S-45	UttarPradesh	Chausa	UC	88.490	0.375	0.485	5.450	0.715	4.205	4.485	15.000	44.105
S-46	UttarPradesh	Chausa	UC	88.285	0.38	0.515	6.050	0.695	4.235	4.075	14.000	45.135
S-47	Punjab	Safeda	PS	81.180	0.425	0.305	6.350	0.605	4.220	11.135	12.000	72.685
S-48	Punjab	Safeda	PS	78.015	0.315	0.515	7.150	0.695	4.250	13.310	15.000	86.475
S-49	UttarPradesh	Langra	UL	76.865	0.365	0.605	7.250	0.635	4.240	14.280	12.000	91.565
S-50	UttarPradesh	Langra	UL	88.350	0.425	0.530	5.950	0.695	4.205	4.050	14.000	44.77

S-51	UttarPradesh	Fazli	UF	81.065	0.325	0.545	6.150	0.705	4.085	11.210	14.000	74.345
S-52	UttarPradesh	Fazli	UF	86.875	0.265	0.465	5.350	0.715	4.205	6.330	15.000	50.905
S-53	UttarPradesh	Fazli	UF	80.980	0.545	0.365	4.350	0.625	4.235	13.135	15.000	73.225

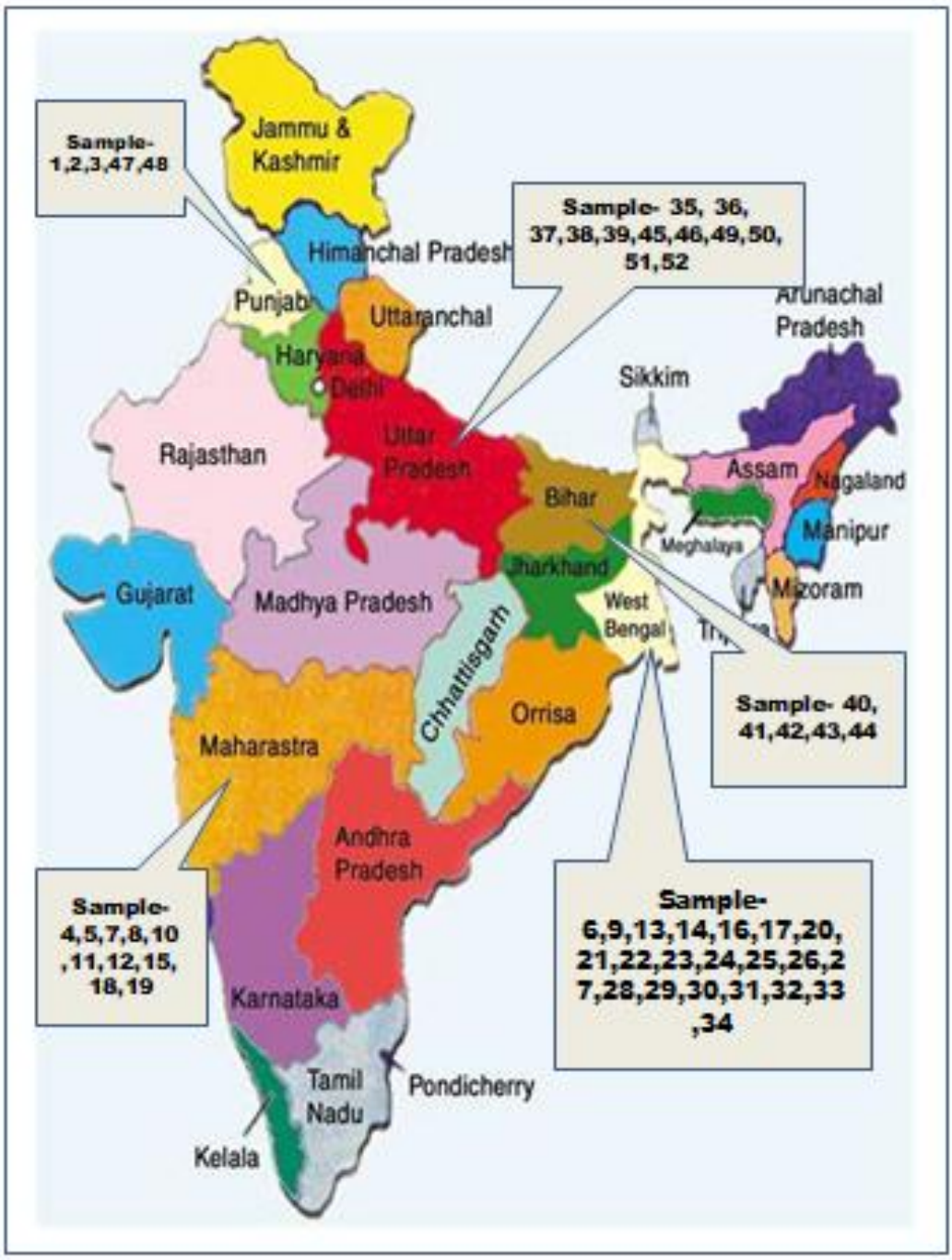


Figure 3. 2 State wise collected Indian mango samples

has been packed in zip lock bags to keep the moisture of the sample constants and stored in a refrigerator with 4 $^{\circ}$C for further analysis.

Figure 3. 3 Mango pulp after extraction from intact fruit

Figure 3. 4 Mango pulp sample at storage in zip lock bags

3.3.3. Chemical Parameter

Proximate analysis were done to measure the different biochemical properties, mainly ash content, moisture content and crude fat content of mango fruit according to the Association of Official Analytical Chemists (AOAC) methods [1]. Experiments were

performed in triplicate to reduce the experimental and manual errors. For the chemical analysis the following chemicals has been used (Table 3.3)

Table 3. 3 Chemical used for the analysis

Sr. No.	Chemicals	CAS No.	Source
1.	Petroleum ether	8032-32-4	Loba Chemie Pvt. Ltd. India
2.	Potassium Sulphate	7778-80-5	Loba Chemie Pvt. Ltd. India
3.	Copper sulphate	7758-98-7	Loba Chemie Pvt. Ltd. India
4.	Sulphuric acid (98% concentrated)	7664-93-9	Loba Chemie Pvt. Ltd. India
5.	40% Sodium hydroxide	1310-73-2	Loba Chemie Pvt. Ltd. India
6.	4% Boric acid solution	10043-35-3	Loba Chemie Pvt. Ltd. India
7.	0.1N Hydrochloric acid	7647-01-0	Loba Chemie Pvt. Ltd. India
8.	Methyl Red Indicator (0.1%)	493-52-7	Loba Chemie Pvt. Ltd. India
9.	1.25% Sulphuric acid	7664-93-9	Loba Chemie Pvt. Ltd. India
10.	1.25% Sodium hydroxide	1310-73-2	Loba Chemie Pvt. Ltd. India

a. Moisture Content

Moisture content was determined using the standard hot air oven (Figure 3.5) AOAC 930.15 method [1].

To measure the moisture content of mango pulp there are five steps that is

- **Petri Dish Preparation:**

  First of all petri dish is dried in which sample was taken for the further analysis. For drying the clean glass petri dish with 120 mm diameter was placed into an oven at 105 $^{\circ}$C temperature for 20 minutes. This step is important to remove the moisture of petri dish. Drying plate is necessary because the moisture of the dish may give a false result. After drying the petri dish was cooled in desiccator.

- **Sample Preparation:**

  After this the sample measurement was taken into the steps. Calibration status of the balance meter was checked before weighting of the sample. Pre-heated moisture free

petri dish was taken to measure the 5 gram of the mango pup sample. Before that the weight of the petri dish was noted down. And then tare the dish weight to measure the weight of the homogenised pulp properly. The 5gm of mango pulp was taken in dish for the test (pre-weight) with spatula. The sample was spread properly on surface of dish by shaking the petri dish.

- **Drying of Sample on Hot Air Oven:**

  The petri dish with sample was placed inside the hot air oven carefully. The sample was placed into it at 150 $^{\circ}$C for 2 hours in a hot air oven. After heating the sample for 2 hours petri dish was removed from oven and transferred to desiccator for cooling.

- **Final Weight:**

  The final weight of the petri dish containing dried sample was recorded and noted down.

- **Calculation:**

  The loss in weight was directed as a measure of moisture content and was determined using the below mentioned formula that is,

$$\text{Moisture (\%)} = \frac{\text{Weight of fresh sample} - \text{weight of dry sample}}{\text{weight of fresh sample}} \text{ x } 100 \qquad (9)$$

Figure 3. 5 Hot air oven for moisture content analysis

b. Ash Content

Ash content was estimated using the standard muffle furnace (Figure 3.6) AOAC 942.05 method [1].

To measure the ash content of mango pulp there are five steps those are

- **Crucible Preparation:**

  A clean porcelain crucible was taken and was placed into oven to remove the moisture content. This step is important to avoid the error in results. The crucible and its lid were kept separately into the oven at 105 $^{\circ}$C for 20 minutes.

- **Sample Preparation:**

  After that the crucible was removed from the oven and transferred to the desiccator for cooling. To get the accurate result the sample weight must be accurate. Before weighting the calibration of the balance machine was checked. After that the crucible was placed on the balance machine and the weight of crucible was noted down and made it tare. Sample weight process should be very faster to avoid the absorption of the external moisture by the sample. Then around 5gm of homogenised mango pulp was taken in porcelain crucible (pre-heated) for the test.

- **Combustion or Burning of The Sample:**

  Sample weight was noted down. After weighting the sample was covered with the lid of the crucible. Finally the crucible was placed into muffle furnace at 525 $^{\circ}$C for about 4 to 5 hours. After finishing the ashing the sample was kept into the furnace till the temperature of the furnace drop below 250 $^{\circ}$C. Crucible was transferred desiccator for cooling.

- **Final Weight:**

  After that the crucible containing with whitish brown ash was placed on the balance machine for taking the weight. To ensure completion of ashing, the method was repeated for half an hour more. This was repeated consequently till the weight became constant.

- **Calculation:**

The ash content present in the mango pulp was calculated with the below mentioned formula that is,

$$\text{Ash (\%)} = \frac{\text{Weight of after ashing}}{\text{weight of fresh sample taken}} \text{ x } 100 \qquad (10)$$

Figure 3. 6 Muffle furnace for ash content analysis

c. Crude Fat

The crude fat of homogenised mango pulp sample was determined with standard AOAC 2003.05 method by solvent extraction using SocsPlus system (Fig. 3.7) [1]. The 5gram of mango pulp was weighed into the fat free cellulose thimbles, and they were placed beneath the SocsPlus condensers using the thimble support clamps. 50 ml of petroleum ether was refluxed over the sample for 1 hour after which it was drained and removed by evaporation for 30 mints. The residual mass in the silica reflux cups was designated crude fat.

Figure 3. 7 SocsPlus apparatus for crude fat analysis

d. Crude Protein

Determination of crude protein content was carried out with standard AOAC 2001.11 Kjeldahl method. Crude Protein analysis was taken by following four major steps of experimentations; those are digestion; distillation; titration and final calculation.

- **Digestion:**

First of all digestion was taking place. In this step the mango pulp was digested by concentrated Sulphuric acid. In this step the presence of the catalyst is important. In this experiment as catalyst a mixture of Potassium sulfate (K_2SO_4) and anhydrous Copper sulfate ($CuSO_4$) in 10:1 ration was used. Before taking the weight of sample the calibration of balance machine was checked properly to eliminate the

experimentation error. Clean weight boat was placed on the balance machine and tare the weight of this. Homogenized mango pulp sample has been transferred into the boat quickly to avoid the external moisture absorption. Around 2 gram of pulp was taken and noted down the sample weight. A kjeldahl flask was taken and marked with particular sample number. Sample was poured into the kjeldahl flask and also 1 gram of catalyst was mixed with it. 20 ml of Concentrated Sulphuric acid was measured with a pipette and poured the acid into the kjeldahl flask. The flask was shacked properly to mix the acid with sample and catalyst. After this, kjeldahl flask was placed on the digestion unit (Figure 3.8) carefully and temperature of digestion unit was set at 230 $^{\circ}$C temperature for 2 hours. Water circulation unit was started at the same time to maintain the temperature of the system. Whenever, the solution was turned into grinish blue, it was indicated that digestion was completed and in this process nitrogen was transformed to ammonium sulfate by hot digestion.

Figure 3. 8 Digestion unit for crude protein analysis

- **Distillation:**

Before going to the distillation the sample was diluted with distilled water. Approximately 20 ml of distilled water was added into the flask with the sample and

shacked properly. Now the sample was poured into a 100 ml volumetric flask using funnel. The same process was done three times to ensure that no digested sample residue was left to transfer from the kjeldahl flask. Sufficient amount of water was added to make the final volume 100 ml. A measuring cylinder was taken to measure 30 ml of 4% boric acid. The 30 ml boric acid solution was poured into conical flask and this flask was placed on the distillation unit where the distillate was collected. The distillation unit (Figure 3.9) was set for experimentation. 10 ml of digested solution was taken from the 100 ml volumetric flask and was transferred into distillation flask and also 50 ml of 40% of Sodium hydroxide (NaOH) and 50 ml of distilled water was added into it. After that the set up was started after setting the temperature at 200° °C. After one hour, 100 ml of distillate was collected. Ammonia was pushed into a known volume of boric acid solution after being liberated by distillation while being exposed to sodium hydroxide (40%) in the process.

Figure 3. 9 Distillation unit for crude protein analysis

- **Titration:**

  For titration 0.1N Hydrochloric acid (HCl) was taken into a burette and noted down the initial burette reading. Few drops of methyl red indicator were added into the conical flask that contained with collected distillate. The color of the distillate was changed from transparent to yellow. This solution was titrated with the 0.1N HCl. When the color of the solution was changed from yellow to orange the titration was completed. Final burette reading was noted down. The percent of protein was estimated from percent of nitrogen as follows:

$$\text{N (\%)} = (V_{HCl}\ X\ N_{HCl}\ )/\text{Sample weight}\ X\ 14\ X\ 100 \qquad (11)$$

N is the normality of the HCL utilized, and 14.00 is the molecular weight of nitrogen. V is the volume of HCL in litres consumed to the end point of titration. Utilizing the proper conversion factor, the percent of nitrogen is changed to the percent of protein (% Protein =% N x 6.25).

e. Crude Fibre

Crude fiber content was determined using the standard acid alkali digestion (Figure 3.10) AOAC 978.10 method. Before the crude fiber analysis 1.25% $H_2SO_4$and 1.25% NaOH was prepared.

- **Preparation of 1.25% H_2SO_4 Solution:**

To prepare the 1.25% H_2SO_4at first a 500 ml volumetric flask was taken and marked. Firstly, 400 ml distilled water was poured into the volumetric flask and 3.5 ml of concentrated Sulphuric acid (98%) was transferred with pipette into the flask carefully. Also, the pipette was washed properly to ensure that no residue of Sulphuric acid was left on inner surface of pipette while transferred into the flask. The flask was rotated several times to ensure the proper mixing of acid with water. Finally enough amount of distilled water was added to that solution to make it 500 ml finally.

- **Preparation of 1.25% NaOH Solution:**

To prepare 1.25% NaOH at first 6.25 gram NaOH pellets was taken. After this a 500 ml volumetric flask was taken and marked. The flask was filled with 400 ml distilled water

and 6.25 gram NaOH pellets was transferred into it. The flask was shacked properly to dissolve the pellets properly in distilled water. After some time mixed water into the solution to make it total volume of 500 ml.

Crude fiber analysis was taken by following steps five major steps of experimentations; those are

- **Boiling of Sample in Acid:**

  200 ml of 1.25% Sulphuric acid solution was taken into a cylinder flask and transferred it into a conical flask. Approximately 2 grams of homogenized mango pulp sample was taken. Before taking the weight of the sample the calibration of the balance machine was checked to eliminate the measurement error. Then the sample was added to conical flask contained with Sulphuric acid and mixed properly and 200 ml of 1.25% H_2SO_4 were heated for 30 min and filtered with a Buchner funnel. The residue was washed with distilled water until it was acid free.

- **Boiling of Sample in Base:**

  The residue was boiled in 200 ml of 1.25% NaOH for 30 minutes, filtered, and repeatedly rinsed with distilled water until it became alkaline-free. Then it was rinsed once with 10% HCl, twice with ethanol, and three times with petroleum ether. In order to completely eliminate the NaOH residue, the residue was lastly rinsed with hot water. The final filtrate was then gathered in a clean, dried crucible until none remained.

- **Drying of the Fiber:**

  Crucible with sample was placed on a hot plate to evaporate the excess water. The residue was put in a crucible and dried at 105 $^{\circ}$C in an oven overnight.

- **Incineration of Fiber:**

  After cooling in a desiccator, crucible was placed in a muffle furnace at 550 $^{\circ}$C for 90 min to obtain the weight of the ash. Again after ashing the crucible was kept inside the desiccator. Weight of the final ash was noted down,

- **Final Calculation:**

  Crude fiber was determined with mathematical formula followed,

$$\text{Crude fiber} = \frac{Weight\ of\ crucible - weight\ of\ crucible\ with\ ash}{Weight\ of\ sample} \text{ X } 100 \qquad (12)$$

Figure 3. 10 FiberoPlus apparatus for crude fiber analysis

f. pH Content

pH content of mango pulp was taken by following two major steps, one is calibration of the pH meter and the final reading.

- **Calibration of pH Meter:**

  Firstly, the shipping tube from the pH meter was removed for cleaning and washing. The probe was washed with distilled water and was cleaned with tissue paper. The probe of the pH meter was inserted in pH buffer 7.0 solution. When the reading of pH meter was reached in 7.0, then CFM button was pressed. After this probe was removed and again was washed and dried with tissue paper. After completely drying, the probe was inserted into the buffer 4.0 solution. When the screen of the pH meter was reached at 4.0 again CFM button was pressed. Again the probe was washed with distilled water and dried properly.

- **Final Reading:**

  The homogenized mango pulp sample was taken into a glass beaker. The probe of the pH meter was inserted into it. pH meter was given an unchanged and steady reading (pH value) of the sample.

Figure 3. 11 pH meter for measurement of pH content

g. Total Soluble Solid Content

The method of determination of the total soluble solid content was followed three major steps. These are refractometer preparation; refractometer calibration and sample addition and final reading.

- **Refractometer Preparation:**

  Refractometer has majorly two parts, one is eye piece and another one is prism. At, prism the sample was kept during the analysis and eye piece was used to determine the ^{0}brix value.

- **Refractometer Calibration:**

  The refractometer was calibrated using distilled water before the sample testing. This step was useful to eliminate all the errors and estimate the minimum error. Few drops

of distilled water were added into the prism of the refractometer. The ^{0}brix value of water was estimated by seeing the readings through eye piece. The reading was in zero. So it was concluded that the ^{0}brix value is zero as no sample was added (only distilled water added).

- **Sample Addition and Reading:**

  Some drops of the sample were added on the prism using dropper. During this step a special care was taken to avoid the formation of the air bubbles on the prism surface. The reading was taken using the eye-piece. The brix percentage was shown in the reading scale of refractometer.

Figure 3. 12 Hand refractometer for measurement of TSS content

h. Carbohydrate

The difference method was used to determine the amount of carbohydrates. Difference method was produced from a mathematical equation, which depicts the subtraction of sum of all other nutritional element from 100 and formula is written as:

$$\text{Carbohydrate (\%)} = 100 - (\%\ \text{Moisture} + \%\ \text{Ash} + \%\ \text{Crude fat} + \%\ \text{Crude fibre} + \%\ \text{Crude protein}) \quad (13)$$

i. Energy Value

Energy value was determined with the equation of Osborne and Voogt and is described as:

Energy (Kcal) = [9 x Crude fat (%) + 4 x Carbohydrate (%) + 4 x Crude protein (%)] (14)

3.3.4. Rheological Parameter

For measuring the viscosity, flow cup method is very popular and was used to determine the viscosity. Basically, the measurement of mango pulp's flow is acknowledged as viscosity, which indicates the consistency and quality of the mango pulp. To measure the viscosity three measure steps was followed,

- **Determination of the Proper Flow Cup:**

Firstly, a viscosity flow cup was taken and calibrate for the measurement. A cup with the estimated flow time of 30 to 100 seconds was selected.

- **Cleaning of the Measuring Device:**

This step is very important in terms of the accuracy of the result. With ethyl alcohol solvent the viscosity cup was properly cleaned and it was made sure that no previous residue was there.

- **Final Measurement:**

After that the flow cup was placed into its stand and was ensured that it was level. To check the level, a glass and a bubble level were placed on top of the cup. The adjustment was done to the feet until a level measurement has been achieved.

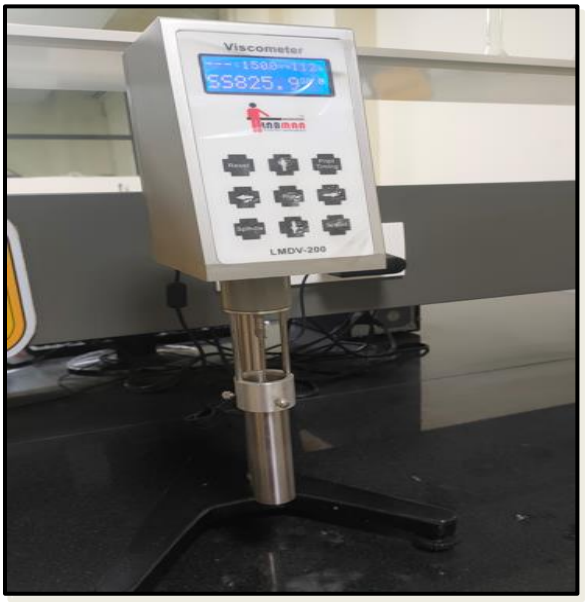


Figure 3. 13 Viscosity cup flow meter for viscosity measurement

3.4. Spectral Parameter

After the physical property analysis the samples has been taken into the CSIR-CSIO Chandigarh for Near Infrared spectra (NIR) collection and UGC DAE Consortium Laboratory for Fourier Transform Infrared spectra (FTIR) collection. The NIR and FTIR spectra data has been collected with NIR FOSS2500 (Figure 3.14) instrument with spectral range of 700-2500 nm and with Perkin Elmer Fourier transform infrared spectrometer (Figure 3.16) with the KBr optical beam splitter and with 8300 to 350 cm^{-1} spectral range.

3.4.1. Instrumentation:

In the year of 1937, April 19, Richard S. Perkin and Charles W. Elmer has been developed an optical instrumental company and named as Perkin-Elmer. He has generated total USD 20000 investment to develop this company. In association of US government they have developed the world's first IR spectrometer, MODEL-12. After them another instrumental company Beckman has produced MODEL IR-1, with some more constructive modifications. After that production IR instrumentation has achieved the popularity and became the most useful tool to produce the precise "fingerprint" of any organic molecule.

a. Near Infrared Spectrometer is having total four components, Figure 3.15 shows the basic instrumentation of the near infrared spectroscopy which includes the source, monochromator, sample holder, transmittance detector or, reflectance detector. This instrument basically works in a particular range (750-2500 nm) of wavelength.

Figure 3. 14 Near infrared spectrometer of CSIR-CSIO Chandigarh laboratory

- **Source:**

As a source of near infrared signal Tungsten-halogen monofilament or NIRED (Near Infrared Emitting diode) can be used.

- **Monochromator:**

Modern gratings are based on interference. That is, they are created by the interference pattern of two lasers impinging on a photosensitive surface, which produces the master gratings from which duplicates are formed (thus, the term holographic). On a reflective (gold, silver, or aluminum) surface, this results in lines with very precise spacing. When a surface receives polychromatic light from a source, it scatters the light in accordance with Bragg's law by acting like thousands of prisms:

$$m\lambda = 2d \operatorname{Sin} \theta \qquad (15)$$

Where, m is the order of diffraction, λ is the wavelength (of light emitted), d is the distance between lines of the grating, and θ is the angle of diffraction.

- **Sample holder:**

Sample holder can be used.

- **Detector:**

The two i.e. reflectance and transmittance are two basic instruments involved with NIR spectroscopy.

i. In the reflectance measurements, the light penetrates only (1-4) $\times 10^{-3}$m from the surface of the sample. When the energy of small penetration passes through the sample it will vary than the transmittance technique. The length of the total path of the sample in the transmittance measurement is investigated into spectrum measurements. Due to transmittance technique measurement can be done on large particles. The higher frequency energy (800 to 1400 nm) is more disposed to front surface scattering as compared to the lower frequency energy.

ii. For transmittance measurements, particle size should not be too small, that signals do not get detected by the detectors.

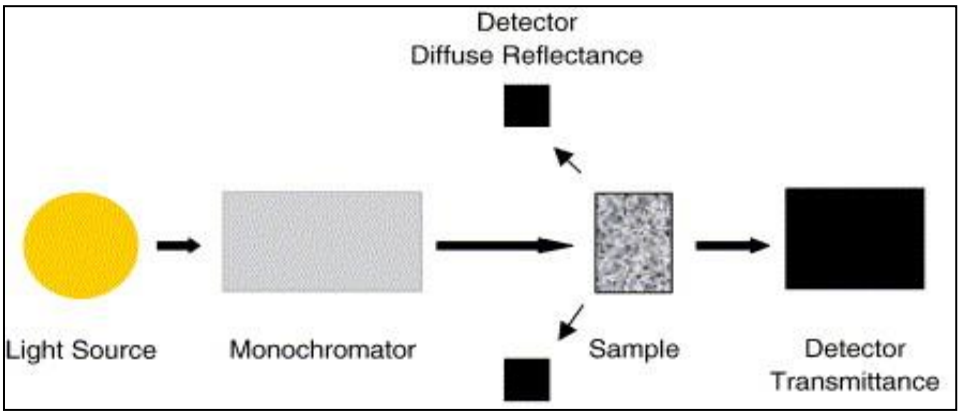


Figure 3. 15 Basic instrumentation of near infrared spectroscopy

Source: [8].

b. Fourier transform infrared spectrometer is having total six different components (Figure 3.17).

- **Source:**

  As a source of infrared signals Global, Nernts glower, Nichrome wire etc. can be used.

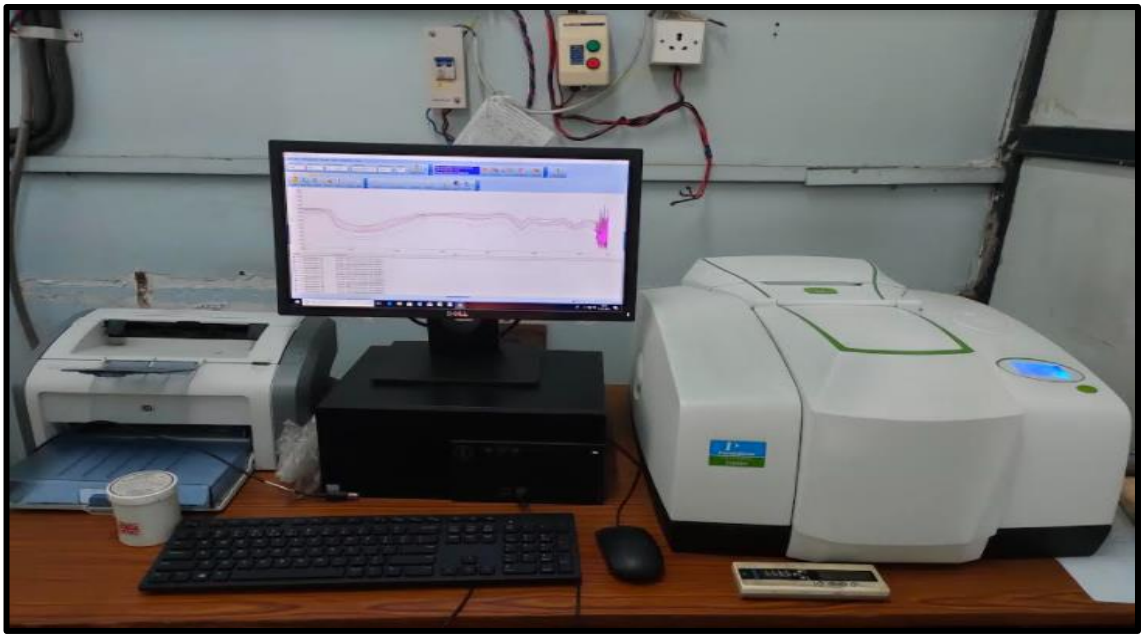


Figure 3. 16 Fourier Transform infrared spectrometer of UGC DAE CSR Center, Kolkata

- **Interferometer:**

  Modulated energy of the source will direct to the interferometer. Each wavelength of an electromagnetic radiation is having a distinct modulation frequency, and by the system it will change into modulated from actual electromagnetic wavelength. Interferometer is having two component mirrors. One of which is moving along with its axis with same velocity and another one splits the coming beam into parts. The optical retardation in a Michelson interferometer is equal to twice the moving mirror displacement. The retardation is zero for all frequencies when the stationary and moving mirrors are equally spaced apart, and the two beams constructively interfere. Therefore, ignoring losses, all of the source energy now arrives at the detector.

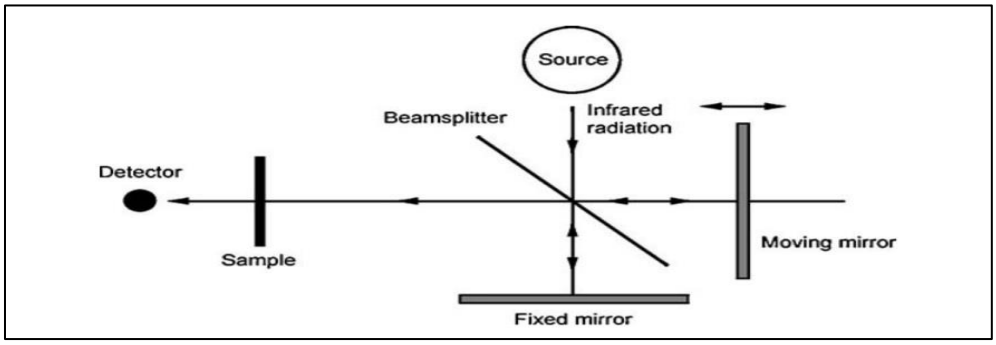


Figure 3. 17 Basic instrumentation of FTIR
Source: [9].

- **Beam Splitter:**

  The NIR energy is partially reflected onto the stationary mirror by the beam splitter and transmitted to the moving mirror with the remaining energy. The beam-splitter recombines the reflected beams and directs them outward. The beam-splitter is layered coatings placed on a substrate. FTIR spectrometer is commonly used for three different substrates that is quartz, CaF_2 and KBr.

- **Laser:**

  To control the wavelength precision, the proper alignment and movement of the mirrors as well as the interferometer the He-Ne laser (monochromatic) is used.

- **Detector:**

There are numerous types of detectors can be used like Deuterated Triglycine Sulfate (DTGS), Mercury Calcium Telluride (MCT), Polycrystalline lead sulphide (PbSe) and Polycrystalline lead selenide (PbSe) etc. for FT-NIR interferometer. Through the sample some NIR signal absorbs and some of them reflects or transmits to the detector selectively.

- **Optical Component:**

  Aluminum flat front surface mirror for redirection and parabolic, ellipsoidal mirror for focusing are used as optical component in spectrometer.

3.4. Statistical Analysis

3.4.1. Correlation Analysis

The relationship between two or more variables is the focus of correlation analysis. The correlation analysis is used to determine how closely related the variables under examination are. The correlation index or correlation coefficient is a measure of correlation that condenses the direction and strength of correlation into a single number [2]. There are two types of correlation that is positive correlation and negative correlation. Positive and negative correlations are also known as direct and inverse correlation respectively. The variation of the variables in same direction indicates positive correlation and the variation in opposite direction is defined as negative correlation [3]. The relationship between the variables' causes and effects is not taken into account by correlation analysis. If there is a cause and effect link between the variables in a bivariate distribution, they are likely to vary in sympathy with one another and consequently exhibit a high degree of correlation. To put it another way, correlation always suggests causality. However, the converse is not true. There are several reasons for which there can be high degree of correlation between two variables that is, mutual dependence, pure chance and both the variables being influenced by the same external factor. Present study has been followed Karl The mathematical method to evaluate the strength or size of a linear relationship between two variables is known as the Pearson's coefficient of correlation method. This method was suggested by Karl Pearson, a great British Bio-metrician and statistician [4, 5].

The correlation among all the biochemical parameters has been examined and few of them are significantly correlated. Correlation analysis and linear regression equation has been determined with the IBM SPSS Statistics version 22.0 software. Principal Component Analysis has been carried with the Clustvis 2.0 (https://biit.cs.ut.ee/clustvis/, accessed by 20 January, 2022) web tool for visualizing clustering of multivariate data.

3.4.2. Principal Component Analysis

Principle component analysis is also known as a dimension reduction method and it can reduce the large number of variables with small number of variables maintain the same information [6]. The principle component analysis is a mathematical tool which can apply to transform a number of correlated variables to uncorrelated variables. Those variables are known as principle components. This method is similar to the multivariate factor analysis and generally it applies on the square symmetric matrix [7]. A linear combination of observed variables is expressed as principal components. The main purpose of this analysis is to collect the correct information and ignore the outliers [8].

3.4.3. Chemometric Methods

The measurement of different fruit qualities (external and internal attributes) can possible with the spectra of visual and infrared region specifically. Multispectral or hyper spectral technologies have been developed to carry out the determination. Organic fruit samples are having some functional groups like O-H, N-H and C-H bond and those are IR active organic moles. IR active molecules can absorb the IR radiations and the absorption spectra can predict about the qualitative parameters. Multispectral imagining can obtain the image of the fruit and determine the different spectral data, while spectroscopy can analyze the optical information of those spectra and determine the physical and biochemical parameters. As, the spectrum obtained after the absorption of organic functional groups are having the complex spectral structures with overlapped combination and overtones and the low signal to noise ratio, generally the visual inspection of the spectra cannot be possible initially. So the elimination of instrumental noise, light scattering and the overlapping of spectral crust and trough is important to get the spectra in the simpler way. For this, several pre-processing techniques can be used [9,

10].

Pre-treatment Methods

There are different pre-treatment methods to do more accurate and perfect the raw spectral data. Among those smoothing, Multiplicative scattered correction, computation of first and second order derivatives are useful in research for the non-destructive quality evaluation of mangoes. Smoothing is a useful pre-treatment technique to eliminate the instrumental noise (high frequency) and improve the signal to noise ratio of a raw spectrum. It is a simpler approach for optimal estimation of the wavelength values by fitting or averaging in a window. If the window is narrow, the spectral resolution should be high. But, it is tough job to determine the window width properly. Smoothing process can be taken into picture through some fitting methods like- Moving average smoothing method, Gaussian filter smoothing, Median filter smoothing and Savitzky-Golay smoothing. Smoothing has also used with the combination of Multiple Scatter Correction (MSC). MSC pre-treatment method was first introduced in 1988 [11]. Due to the different particle size, physical aspects and non-homogeneous distribution within the raw spectral data, non-uniformity comes and introduces the term light scattering of the spectra. As a solution of this MSC is being the best method by linearized every spectrum to an ideal spectrum. The average spectrum of the calibration sets are known as ideal spectrum and are fitted with every spectrum through least square fitting. When the relationship between the sample concentration and the absorbance of wavelength is good, then MSC is considered to be a useful way. Drifting and the light scattering can also be removed by the derivative correction (first and second order derivative) pre-treatment approach. This approach has a permanent solution and widely used spectral resolution enhancement, distinguish the overlapped crust and trough of the wavelength peaks and remove the background interference. Sensitivity of the spectral data can also be increased through those pre-treatment techniques. Directly finite differences and Savitzky-Golay derivatives has generally used as derivative correction methods. Smoothing is important parameter before the derivatization of the spectral data. Some literature has claimed that Savitzky-Golay derivative correction method is feasible and useful, whereas some depicts

about the second order derivative the same.

Other than that there are more useful tools for pre-treatment techniques such as offset correction (method for subtracting the first few wavelengths from each spectrum, just to manipulate the baseline drift), De trending (also useful to remove the baseline drift), Standard Normal Variant (remove deviation due to different particle size and the scattering. This method is more suitable than MSC to eliminate the errors), Orthogonal signal Correction and Wavelet Transformation (naturally used for data compressing, filtering and smoothing. Daubechies and Symlets are the functions which are more suitable and effective among all wavelet functions [10].

Variable Selection Methods

Chemometric calculation has used the wavelengths as the variables. The purpose of wavelength extraction from the whole spectral range is to simplify and compress the spectral data for the model development and accurate prediction. Those extracted particular wavelengths for appropriate prediction are known as characteristic wavelengths (CW). Another purpose for variable selection is to reduce the time of calibration process. The calibration model can be more stable; model complexity can be reduced for the economical purpose and decreased the model error by selecting the characteristic variables. For quality measurements the equation can be expressed like this for the wavelength selection:

$$Y = B_0 + \sum B_i X_i \quad (16)$$

Where, Y is dependent parameter, B_i are the weighed regression coefficients and X_i are the spectral variables [12].

There are many methods to the spectral data from the whole spectral range. Regression Coefficient (RC), Successive Projection Algorithm (SPA), Genetic Algorithm (GA), Competitive Adaptive Reweighted Sampling (CARS) and Loading Weighs (LW) are the most commonly used approaches. The values of the Regression Coefficient can depict the performance of the calibration model developed after the partial least square regression. Each wavelength of the prediction model has the regression coefficient value associated with this and those values can predict the characteristics wavelengths. SPA is a forward

selection method and follows the iterative method to project a value of new wavelength started from one wavelength till a certain number has reached. SPA is considered to a highly performed extraction method which can dominate the characteristics wavelengths (wavelengths having the maximum information about the sample) and eliminate the overlapped wavelengths. While making the calibration model with the regression analysis loading weighs are key indicator which can take out the accurate variables (wavelengths) from the whole range of spectral data. The wavelengths having highest value of loading weighs consider being the characteristics wavelengths. In this case the number of latent variable is exactly same as of characteristics wavelengths. The finest value of the wavelength should have the minimum root mean square error for cross validation. Competitive Adaptive Reweighted sampling is a proper technique which can accumulate the characteristics wavelengths from the range of data through the adaptive reweighted sampling method. A calibration models have serial of variables and have cross validated also and this variable selection can be done through CARS with best compliment. Another effective approach is Generic Algorithm technique, which is based on the fitness function. Genetic operation along with the crossover, selection and mutation GA attains the finest result. It is an iterative method and having at least 100 sets of data and for the fitness function, the root mean square value has used. Reduction of the numbers of spectra can be done through irrelevant spectra elimination and the variables with the highest values of frequency during iteration are the characteristic wavelengths [10].

Discriminant Methods

Pre-processed data with extracted characteristic variables are final inputs for the development of the calibration model. For developed the model some Principle Component Analysis (PCA), Support Vector Machine (SVM), Linear Discriminant Analysis (LDA) and Partial Least Square Discriminant Analysis (PLS-DA) are some effective discriminant applications. PCA is an analysis method has determined the values of the principle component sets. Principle components are the combination of original but uncorrelated spectral data. Principle components are associated with the data variance and this approach always incorporate with the other discriminant techniques. Support

vector Machine method is useful for the linear as well as nonlinear both the data. Kernel function is the modulator in this case which can classify the data well. This method mapped the classified data from the original space to feature space and create a hyper plane for proper discrimination. Linear Discriminant Analysis has usefulness in machine learning broadly. It separates different classes of the object and works on vector space. It transform the linear data from n-dimensional to m- dimensional vector space (where n>m). Partial Least Square Discriminant analysis carried out with the PLS regression method. It deals with the dummy variables (numeric variable that represents the categorical data) and the final variables of the PLS component are decided after the cross-validation [10].

Calibration Methods

Several multivariate calibration analysis methods can be developed with the particular selected wavelengths to predict the different quality parameter. To calibrate the data pre-processing is important. With the pre-processed data different analysis techniques can be applied for the prediction. Mainly three methods used broadly for the research i.e. Partial Least square (PLS) regression, Multiple Linear Regression (MLR) and Principle Component Regression (PCR). Multiple Linear Regression measures the value of independent parameter after the linear combination of the spectral data at each point of the wavelength. This method has decayed the efficiency due to multicollinearity between the variables. Although, the difference between the values of correlation coefficient of prediction (Rp) and correlation coefficient of calibration (Rc) has too large, which is a major drawback in this calibration method. On another note, the other method like Principle Component Regression can eliminate the problem of the multicollinearity between the variables or principle components. Principle components have used as the replacement of dependent spectral variables and predict to fit a MLR calibration model. Though, this method has solved the major problem of MLR analysis, but PCR is having one more major gap. Those principle components have carried very tiny information about the dependent variable. Finally, the PLS regression method have developed a better way for calibration and overcame all the drawbacks of the MLR and PCR analysis. This

method has selected more precise orthogonal data set (latent variable) from the whole range for the betterment and more accuracy. It can omit the problem of multicollinearity and the number of considered latent variable is less here. For the NIR calibration instead of those above methods; Artificial Neural Network (ANN) has widely received an importance. It has three different layers (Input, Hidden and Output) of neurons within it. The neuron present in each layer has an interconnection with each other and connected with weight factor. Non-linear function has used to determine the value of neuron present in hidden layer and that is equal to the weighted sum of the values of neurons of the input layer. Similarly in case of the output layer neuron the weighted sum value of the hidden layer's neuron has taken. Although ANN has effective potentiality to calculate the accurate predictive values of the calibrated parameters but it has some major drawbacks. It faces some difficulties when visualize, it runs with very slow training speed and also has some problem of over fitting [10].

Model Evaluation

This is the final step to get a calibration model to predict the quality attributes. After calibrate the model it is very essential step to check the prediction efficiency to evaluate the model's accuracy [10]. Bias is defined as the average deviation between the reference value (X_n) and the predicted value (Y_n) of the validation set. Bias value should be zero for a good model [13].

$$\text{Bias} = \frac{1}{n}\sum(Xn - Yn) \tag{17}$$

The characteristics of a good model should have the correlation coefficient (R) and the minimum value of the standard error of calibration (SEC) and standard error of validation (SEV). The coefficient of correlation (R) can be formulated as [13]

$$R = \frac{\sum(Xn-Xn')(Yn-Yn')}{\sqrt{(Xn-Xn')2(Yn-Yn')2}} \tag{18}$$

And the coefficient of determination (R^2) is formulated as [13]

$$R^2 = \left(\frac{\sum(Xn-Xn')(Yn-Yn')}{\sqrt{(Xn-Xn')2(Yn-Yn')2}}\right)^2 \tag{19}$$

It is also notable that there is should be very little difference between the SEC and SEV values for excellent performance of the model. If the difference is more then it predicts

the presence of large number of latent variables and principle components in case of Partial least square regression and Principle component regression method [14]. The standard error of calibration set (SEC) can be defined as the deviation of the differences between the reference value (X_n) and the predicted value (Y_n) of calibration set and can be formulated as [13]

$$SEC = \sqrt{\frac{1}{n}\sum(Xn - Yn - Bias)^2} \quad (20)$$

And the standard error of validation set (SEV) can be defined as the standard deviation of the differences between the reference values (X_n) and the predicted value (Y_n) of the validation set and can be formulated as

$$SEV = \sqrt{\frac{1}{n}\sum(Xn - Yn - Bias)^2} \quad (21)$$

On other hand some more parameters like RMSEP and RMSECV and residual predictive deviation (RDP) are having also importance to validate the model's accuracy [10].

References

1. AOAC. (2000). Official Methods of Analysis of the Association of Official Agricultural Chemists, Washington, DC, pp-125-139.
2. N. J. Gogtay and U. M. Thatte, J. Assoc. Physic. India, **65**, 78-81 (2017).
3. N. B. Chapman and J. Shorter, *Correlation Analysis in Chemistry: Recent Advance* (Plenum Press, New York and London, 2012), pp. 357-390.
4. S. C. Gupta, *Fundamentals of Statistics* (Himalaya Publishing House, 2018).
5. C. R. Kothari and G. Garg, *Research Methodology Methods and Techniques, (*New Age International Publishers, 2019).
6. I. Jolliffe, *J. Multivar. Anal*., **188**, 104820 (2022).
7. H. Abdi and L. J. Williams, *Wiley Interdiscip. Rev. Comput. Stat*., **2**, 433-459 (2010).
8. D. Granato, J. S. Santos, G. B. Escher, B. L. Ferreira and R. M. Maggio, *Trends Food Sci. Technol*., **72**, 83-90 (2018).
9. A. A. Munawar, D. von Hörsten, J. K. Wegener, E. Pawelzik and D. Mörlein, *Eng. Agric. Environ. Food*, ***9***, 208-215 (2016).
10. H. Wang, J. Peng, C. Xie, Y. Bao and Y. He, *Sensors*, **15**, 11889-11927 (2015).
11. J. L. Ilari, H. Martens, T. Isaksson, *Appl. Spectrosc*., **42**, 722–728 (1988).
12. U. Kamboj, P. Guha, and S. Mishra., *Anal. Lett.,* **50**, 1754-1766 (2017).
13. Y. A. Purwanto, W. Z. Putri Wulandari, U. Ahmad, M. Sutrisno, Y. Makino, S. Oshita, Y. Kawagoe and S. Kuroki, *Int. J. Eng. Technol.* **13**, 70-73 (2013).
14. S. N. Jha, A. R. P. Kingsly and S. Chopra , *Biosys. Eng.*, **94**, 397-402 (2006)

    .

Chapter-4: Physical, Chemical and Rheological Characteristics of Mango

Chapter-4

Physical, Chemical and Rheological Characteristics of Mango

This chapter focuses on the different physical, chemical and rheological attributes of the mango fruit. The discussions of these properties are important to know about the external and internal quality parameters of mango.

4.1. Introduction

Fruits, as a vast source of essential nutrients, vitamins and minerals, have an active part in human daily diet. Moreover, dietary fiber present in fruits plays a significant role in the case of cardiovascular disease and obesity [1]. In addition to this, fruits can boost up human immunity system without adding many calories. Having all these beneficial supports, Mango has the name and fame as the King of fruits [2, 3]. Mango fruit offers a substantial amount of vitamin A, Vitamin E, Vitamin C and Vitamin B_6. Mango fruit can balance the blood sugar level and can handle the glycemic level at pre-diabetic and diabetic stage. Mango fruit is also known as the procurement of ulcerogenic, neohroprotectant and gastrointestinal diseases [4]. India, as the largest supplier of the mangoes, fulfills approximately 55% of the global demand, followed by China, Thailand, Indonesia and Pakistan. In India, mango production covers an area of nearly 1.23 MHa against an annual quantum of 10.99 MT. It happens to grow up within the tropical and subtropical region up to an altitude of almost 1500 m from sea level [5]. Totapuri, Baganapalli, Kesar, Alphanso, Chausa, Langra are most common among the Indian varieties (Jha et al., 2010). Uttar Pradesh is the top ranking state (23.06%) in terms of mango production within the country, followed by Andhra Pradesh (16.07%), Karnataka (9.29%), Telangana (8.54%) and Bihar (7.52%). Mango has a worldwide economical acceptance due to its flavor, texture and chemical constituents and with the growing food industries the demand of tropical fruits has been increasing day by day [6, 7]. Mango

pulp can be preserved and from this many food products like ice cream, jam, jelly, fruit cocktail, nectar, squashes, mango shakes, mango toffee has been made. Maturity and quality of the mango fruit can be assessed with surface firmness, glossy skin texture and juicy flavor by maximum consumers. But, most of the time this kind of assessment with only eye vision has proved as a mislead [8]. For Indian market mango has a great demand and also it has a worldwide export demand. The need to study of physical attributes of the mango fruits are important to providing valuable engineering data required for designing and to develop the machines, structures and equipment for transporting, which includes handling, transporting , processing and storage of mango fruits. Grading, sorting, packaging and cleaning of the mango fruits before export to various places from the different farms the two physical attributes i.e. shape and size strongly considerable [9, 10]. Other important physical properties like specific gravity and density is relevant to calculate the terminal velocity, thermal diffusivity in heat transfer and Reynolds number for pneumatic and hydraulic handling of products. Storage, transportation and separation system can be categorized with mass of the intact fruit and the bulk density and porosity. The physical parameters have its own mathematical expression for determination [11]. Shape is an important key parameter of mango fruits for the export purpose. Two components that is longitudinal and lateral cross section of the fruit consider as main components for shape determination [11, 12]. For this purpose, the companies concerns have to preserve mango pulp obtained from fresh mangoes [13].

Mango has a worldwide economical acceptance due to its flavor, texture and chemical constituents and with the growing food industries the demand of tropical fruits has been increasing day by day [6, 7]. Mango is one of them, which produces bulk of different food products. Mango pulp can be preserved and from this many food products like ice cream, jam, jelly, fruit cocktail, nectar, squashes, mango shakes, mango toffee has been made. For this purpose, the companies concerns have to preserve mango pulp obtained from fresh mangoes [13].

Mango pulp is a viscous fluid and to know the transport property of the pulp viscosity is the major parameter among all the rheological attributes. Food science and food

processing systems are having a big role of the transport property of the fluid [14].

4.2. Materials and Method

4.2.1. Characterization of Physical Parameters

Fifty one fully ripe mango of different Indigenous varieties has been collected from the different mango farms of five states of India, namely, Bihar, Maharashtra, Punjab, Uttar Pradesh and West Bengal. Collected fruits have been brought to the Advanced Analytical Testing Laboratory, Kolkata, West Bengal for chemical analysis. The fruits are then washed properly by distilled water to remove the heat, external dirt, dust, pollutants and the pesticide residues. Collected fruits have been brought to the Physics Research Laboratory, Lovely Professional University, Punjab, India for the further analysis. The major (a), minor (b) and intermediate (c) axes has been measured with the vernier caliper. The pulp weight, peel weight and kernel weight has been measured with scientific balance meter.

4.2.2. Characterization of Chemical Parameters

The biochemical analysis of the mango homogenized pulp was taken with the AOAC standard method. Crude fiber, crude protein and crude fat have been determined with Acid-alkali digestion method, Kjeldahl method and Soxhlet method respectively [15]. pH content and TSS values have been determined with the pH meter and hand refractometer. Detailed methods have been already discussed in previous chapter (chapter-3). Results have been recorded in the triplicate form and finally the mean value was taken. Energy values have been calculated with the equation developed by Osborne and Voogt and described as,

$$\text{Energy} = [9 \text{ X fat } (\%) + 4 \text{ X carbohydrate } (\%) + 4 \text{ X protein } (\%)] \quad (22)$$

4.2.3. Characterization of Rheological Parameters

In rheological parameter viscosity of the mango pulp has been determined and the detailed method has been already discussed in previous chapter (chapter-3). Cup flow method has been performed to estimate the viscosity of mango pulp.

4.2.4. Correlation Analysis

Present study has been followed Karl Pearson's coefficient of correlation method which

can be defined as the mathematical method to measure the intensity or the magnitude of linear relationship between two variables. This method was suggested by Karl Pearson, a great British Bio-metrician and statistician. Detailed methods have been already discussed in previous chapter (chapter-3). The correlation among all the biochemical parameters has been examined and few of them are significantly correlated. Correlation analysis and linear regression equation has been determined with the IBM SPSS Statistics version 22.0 software.

4.3. Results and Discussions

4.3.1. Determination of Physical Properties

Descriptive statistics of 51 different mango samples of physical attributes of the intact mango fruit has been calculated and shown in the Table 4.1. The different physical parameters like pulp, kernel and peel weight, major, minor and intermediate diameter, size, surface area and sphericity has been determined using the mathematical formulations. Size has been calculated with the formula

$$\text{Size (Dg)} = (a \times b \times c)^{1/3} \quad (23)$$

This attribute is described with three major parameters those are major and minor diameter and the intermediate diameter. In upon expression ‘a’ is the major diameter, ‘b’ is the intermediate diameter and ‘c’ is the minor diameter. Basically those are the spatial dimensions that are length, width and breadth.

Surface area has been determined with the mathematical equation that is

$$\text{Surface area (Sa)} = \pi (Dg)^2 \quad (24)$$

Where, Dg is the geometrical mean diameter. So, it is clearly showing that surface area is directly proportional to square value of the geometrical mean diameter.

Sphericity of the mango fruits can be expressed as ϕ and the defined expression for the sphericity of the mango fruit is given below

$$\Phi = \frac{Dg}{a} \times 100\% \quad (25)$$

Where, Dg is the geometrical mean diameter and ‘a’ is the major diameter.

All the collected data has been determined with the above mathematical formulations and described in the current study with the descriptive statistics. Minimum, maximum and

mean values, standard deviation and coefficient of variance has been calculated and presented in Table 4.1. The minimum and maximum values of the seed weight, pulp weight and peel weight has depicted that the sum of peel and seed weight is less than the pulp weight. It has indicated that mango is a fleshy fruit. Also the range of major, minor and intermediate axis values has been shown very small variations, which means the mangoes collected from all region have almost same sizes.

Table 4. 1 Descriptive statistics of the physical parameters of 51 intact Indian mango samples

Parameter	Mean	Standard Deviation	Coefficient of variance	Minimum	Maximum
Intact weight (gm)	283.1765	104.9849	0.3707	171.0000	731.0000
seed weight (gm)	45.3725	12.5315	0.2762	26.0000	85.0000
Peel weight (gm)	54.8627	19.1594	0.3492	29.0000	114.0000
Pulp Weight (gm)	182.9412	86.4258	0.4724	91.0000	551.0000
Major axis (a) (cm)	10.2445	1.3193	0.1288	7.5300	12.4300
Minor axis (b) (cm)	6.3643	0.8119	0.1276	4.6400	8.7600
Intermediate axis (c) (cm)	7.0478	0.9658	0.1370	5.5800	10.2800
Size (cm)	7.6886	0.7638	0.0993	6.3848	10.1814
Surface area (cm^2)	187.4777	38.7081	0.2065	128.0432	325.6001
Sphericity (%)	0.7578	0.0813	0.1073	0.5834	0.9065

4.3.2. Determination of Chemical Properties

Table 4.2 shows the results of descriptive analysis for the proximate compositions of different mango samples. Descriptive statistics from mean, standard deviation, maximum, minimum value and the coefficient of variance of all the physical and chemical parameters like moisture, ash, crude fat, crude fiber, crude protein, pH, carbohydrate, total soluble solids (TSS) and also chemical energy has been depicted.

Moisture Content

Moisture content depicts the presence of water in any food product. Water content is responsible for the shelf life for the mango fruit and depends upon the different harvesting techniques and growing atmosphere of mangoes. The microbial contamination can be measured with the water index of the mango fruit. Table 4.2 shows the descriptive statistics of the moisture parameter of 51 mango pulp samples and it is clearly showing the mean value of moisture content as 82.5975. Samples have been taken from different parts of India that is Uttar Pradesh, Bihar, West Bengal, Punjab and Maharashtra. Dusheri variety from Punjab is having the highest water content that is 89.115%, whereas the same cultivar from Uttar Pradesh is having lowest amount of water i.e., 76.815%. This variation in the moisture content in mangoes may be caused due to different storage conditions [16]. It is reported that fruits are having mostly 70% to 90% moisture content. In this study every variety is having more than 75% water content, which is very significant. Variation of moisture content with the different sample varieties has been presented in Figure 4.1.

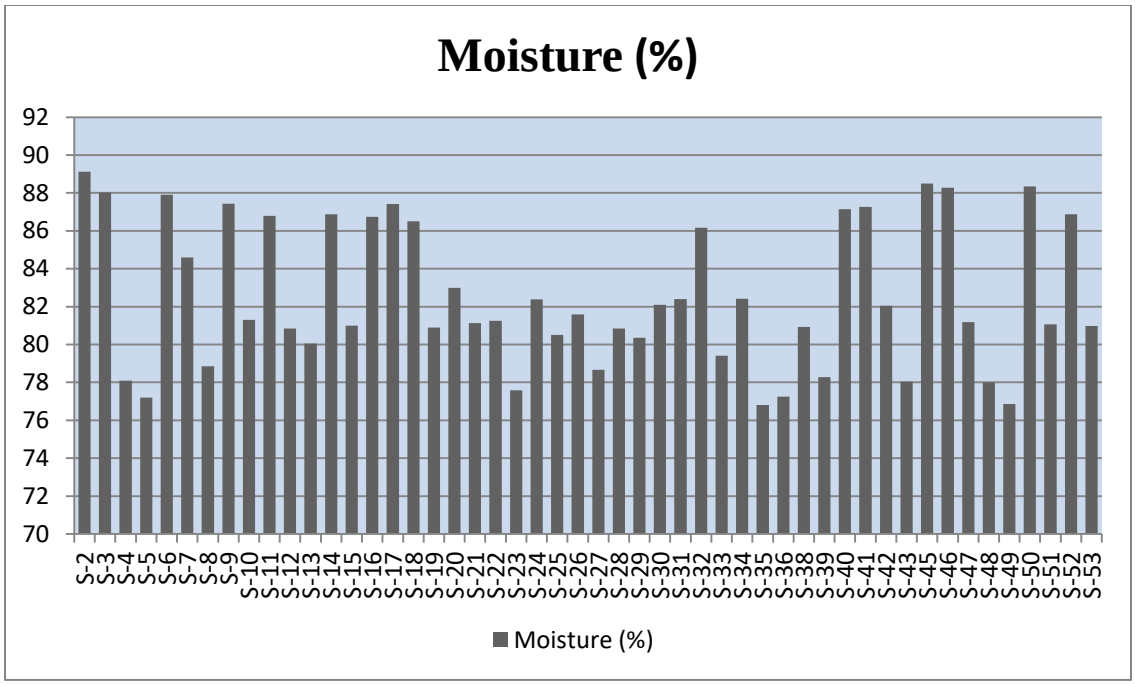


Figure 4. 1 Graphical presentation of the variation of moisture content with the different mango varieties.

Ash Content

Ash content is defined as the total mineral content present in any food sample. Mango is having very less quantity of ash. In this study maximum variants are having the ash values less than 1%, but basically two specific varieties that is Alphanso and Imampasand from Maharashtra and Langra from West Bengal has the ash value greater than 1% that is 2.145%, 2.111% and 7.85% respectively. The highest value observed for the ash content of mango fruit is 7.85% and Dusheri mango of Uttar Pradesh region has the lowest value of ash content that is 0.085%. These maximum and minimum values of the particular varieties might be related to the higher and lower concentration of mineral respectively [1]. Table 4.2 presents the mean value of the ash content that is 0.6273. Variation of ash content with different mango varieties has been shown in Figure 4.2.

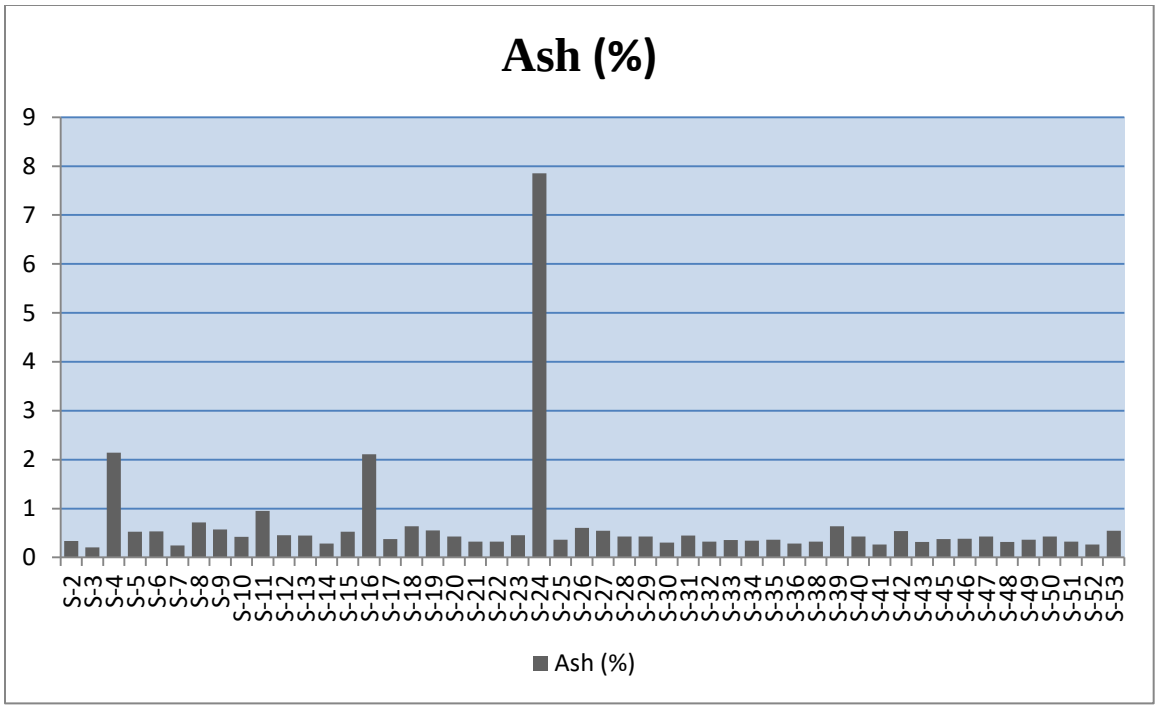


Figure 4. 2 Graphical presentation of the variation of ash content with the different mango varieties.

Crude Fat Content

Crude fat content is considered as a macronutrient component present research shows that mango pulp has very low fat content which provides the benefit of human health. Generally fruits are having the fat content less than 1%. Mean value has been calculated as 0.3716. Highest and lowest crude fat concentration has been observed in Dusheri mango of Uttar Pradesh region and Golapkhas of West Bengal respectively. So, Golapkhas mango may be very good for health as it has very less amount of fat content and can avoid the problem like high pressure, cholesterol, obesity etc. Variation of crude fat content with the mango varieties has been shown in Figure 4.3.

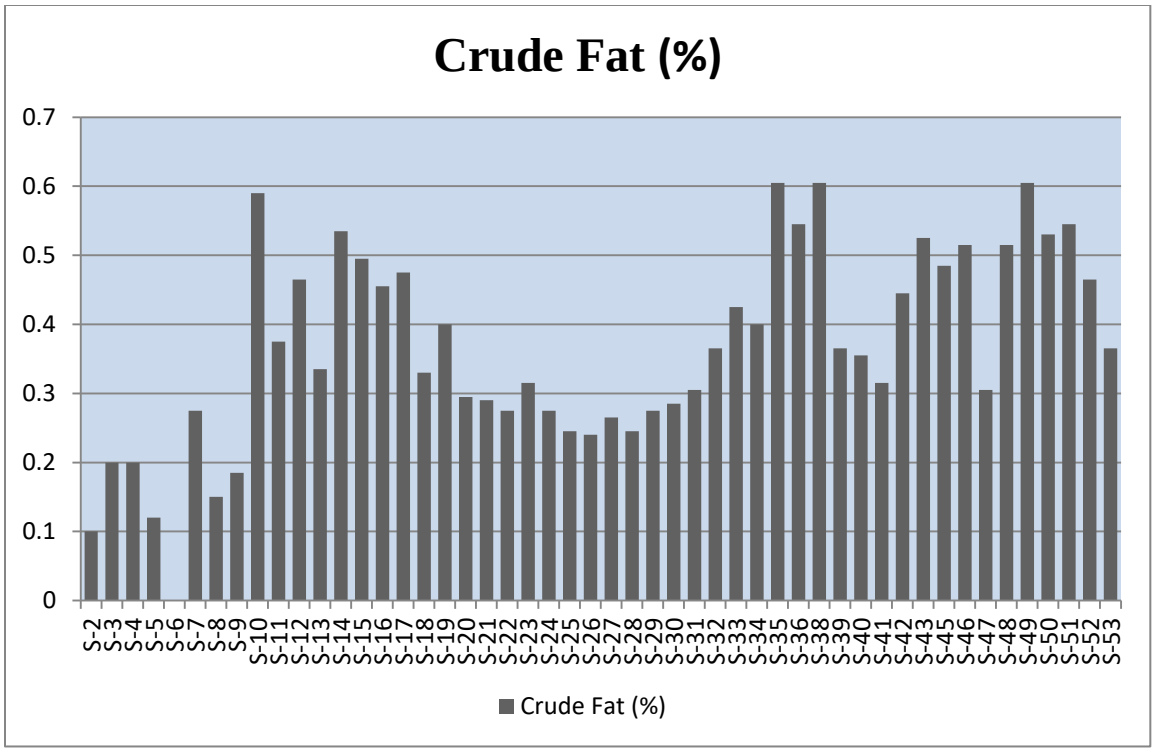


Figure 4. 3 Graphical presentation of the variation of crude fat content with the different mango varieties

Table 4. 2 Descriptive statistics of chemical parameters of 52 different samples of Indian mango using standard analytical methods

Chemical parameters	Mean	Standard Deviation	Maximum	Minimum	Coefficient of variance
Moisture (%)*	82.597	3.777	89.115	76.815	0.045
Ash (%)*	0.627	1.082	7.850	0.085	1.726
Crude Fat (%)*	0.371	0.145	0.605	0.000	0.390
Crude Protein (%)*	5.562	0.817	7.250	3.750	0.146
Crude Fiber (%)*	0.592	0.700	0.715	0.470	0.118
pH	3.975	0.443	4.790	2.740	0.111
Carbohydrate (%)*	10.248	3.679	16.250	2.810	0.359
TSS (%)*	13.278	2.236	17.000	10.000	0.168
Energy (Kcal/100 gm)	66.590	15.7865	91.925	37.915	0.237

* % = gm/100 gm

Crude Protein Content

It has been observed that mango is good in protein content source. Table 4.2. of the descriptive statistics of the chemical compositions of mango has shown that the highest and lowest values of the crude protein that is 7.25% and 3.75% respectively and it also depicts the mean value as 5.5624. Dusheri mango of Uttar Pradesh has the highest and Himsagar mango from West Bengal has the lowest presence of crude protein quantity. The mean value has observed for protein content 5.5624, which is shown in Table 4.2. Variation of crude protein content with the varieties of mango has been presented in Figure 4.4.

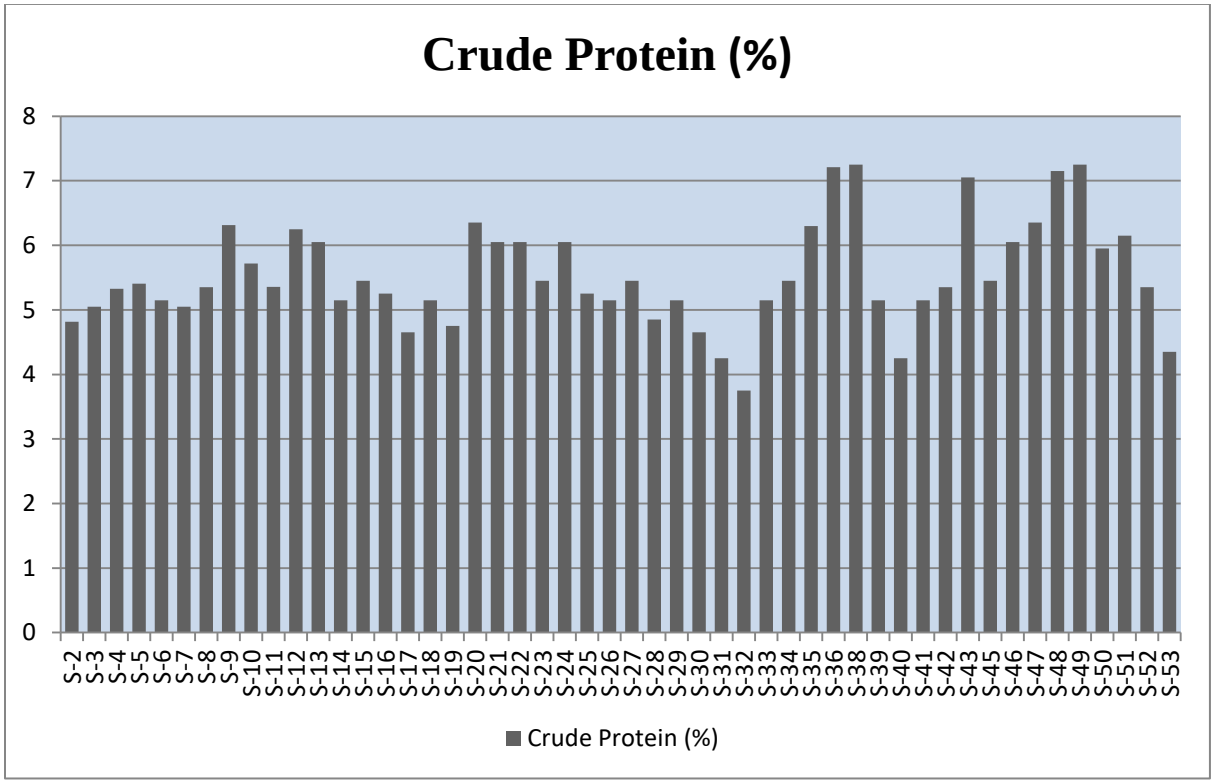


Figure 4. 4 Graphical presentation of the variation of crude protein content with the different mango varieties.

Crude Fiber Content

Mango fruit is a good source of fiber and crude fiber can be defined as the carbohydrate which cannot be digested with our body enzymes. As per the descriptive statistics (Table

4.2) show the highest and lowest value of the crude fiber has been recorded as 0.715% and 0.47% respectively. Mean value of the crude fiber content has been determined as 0.5922308. Imampasand of Maharashtra has the lowest crude fiber concentration whereas the highest value has been recorded for the Fazli mango of West Bengal. In most of the verity it has been observed that the percentage of crude fiber is more than 0.5. Variation of crude fiber content with the mango varieties has been presented in Figure 4.5.

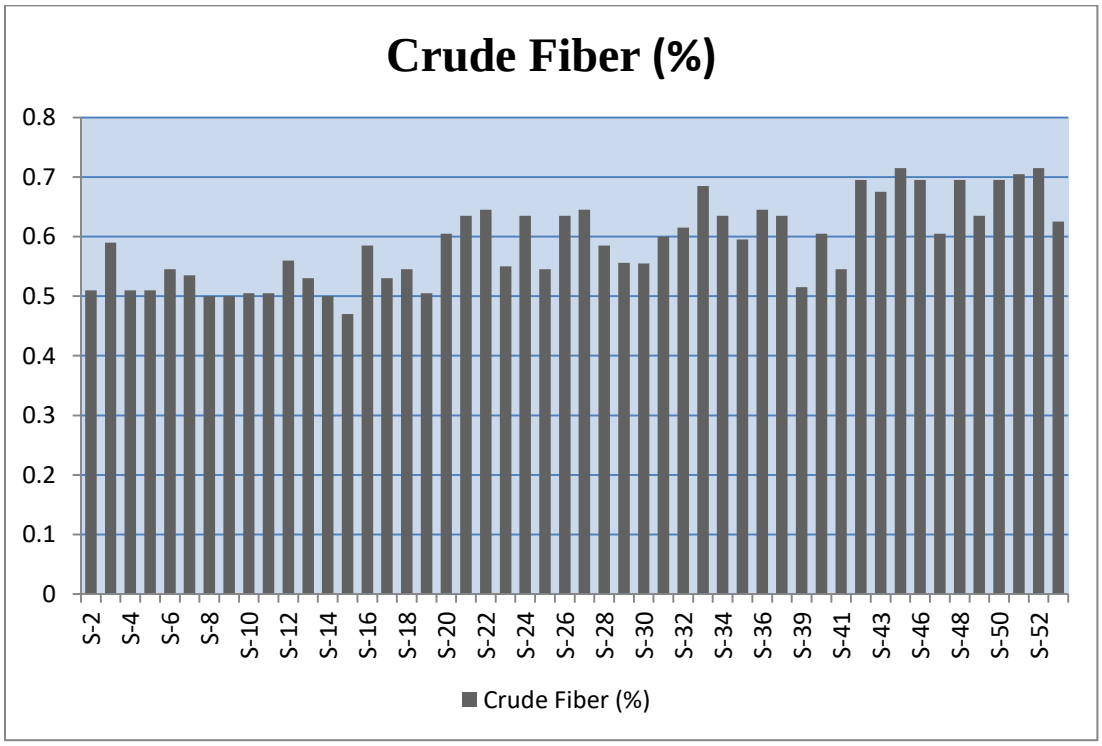


Figure 4. 5 Graphical presentation of the variation of crude fiber content with the different mango varieties.

Carbohydrate and Energy Value

Carbohydrate present in a mango fruit can be characterized as starch and pectin. In this study it has been recorded that the presence of carbohydrate in a mango fruit varies from 10 to 15% in most of the cases. Although the present study has been reported the lowest percentage of carbohydrate content is 2.81%, which is very less as compare to standard value. This value has been observed for the Langra cultivar collected from West Bengal. The highest value for the same is determined as 16.25% for the Imampasand mango of

Maharashtra. Table 4.2 has depicted the mean value of carbohydrate content as 10.248923. The changes of the carbohydrate values with the variation of mango varieties have been presented in Figure 4.6.

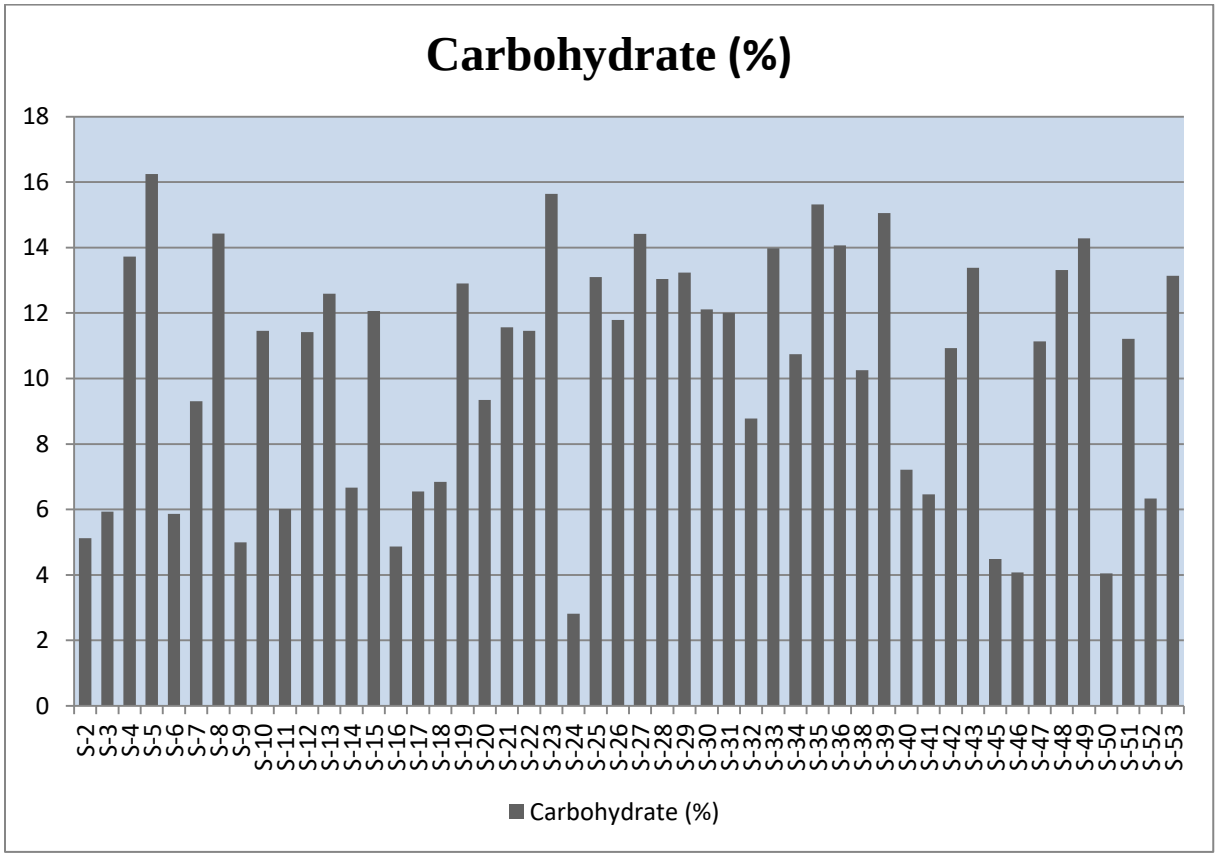


Figure 4. 6 Graphical presentation of the variation of carbohydrate content with the different mango varieties

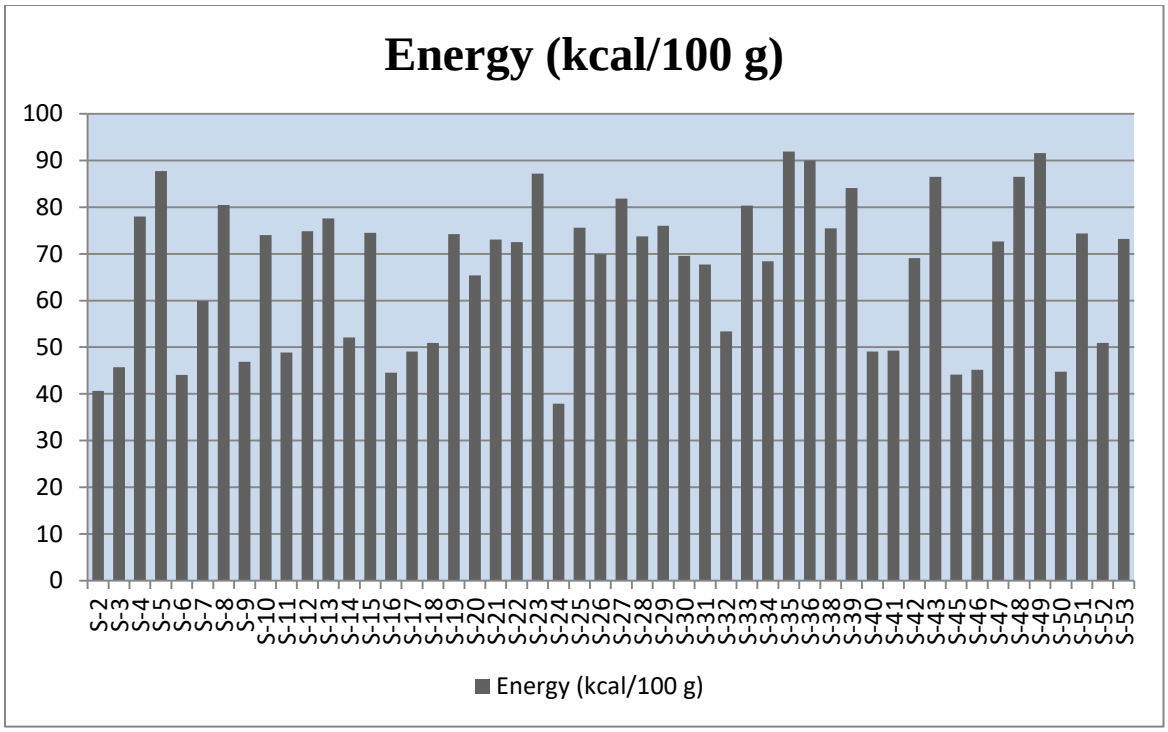


Figure 4. 7 Graphical presentation of the variation of energy values with the different mango varieties

This study gives nutritional and immunity enhancement effect of mangoes, it also builds the energy in body and act as an anti-inflammatory substance. The highest energy any human body can get from the banana fruit, after that mango is in queue. Highest energy value observed in mango fruit is more than 90% that is 91.925 kcal/100 gm and lowest value has been recorded as 37.915 kcal/100 gm. Dusheri mango from Uttar Pradesh is giving the highest value of energy whereas lowest content is giving by one of the cultivar of West Bengal, Langra. Figure 7 and Figure 8 explain the graphical approach that how values of these two parameters change with the different cultivars. The changes of the energy values with the variation of mango varieties have been presented in Figure 4.7.

Total Soluble Solids

Total soluble solid or TSS is defined as the total soluble and insoluble sugar present in the mango and is counted as a major quality parameter of mango fruit. The composition and the texture of mango pulp can be directed with this parameter. This parameter can be changed with storage period of mango with different temperature. Sweetness of mango

can be examined with this property. This parameter generally determined with brix values. The highest and lowest value recorded for the mango fruit is 17 and 10 respectively. There are many varieties which is having same refractive index that is 10^{o}Brix. Dusheri from Punjab, Alphanso from Maharashtra, Surmaiyapally, Golapkhas and Langra from West Bengal has been recorded the same value. The highest value has been observed for the Golapkhas, Bombay Green and Langra from West Bengal. This parameter might be depends on the temperature. The temperature of West Bengal might be higher than other places like Punjab, Bihar, Maharashtra, Uttar Pradesh, which factor have been worked upon the TSS content. On the hand, ripening process and microbial spoilage also affects the TSS content [1]. A graphical data representation has been done in Figure 4.8 that is the variation of moisture content with different varieties.

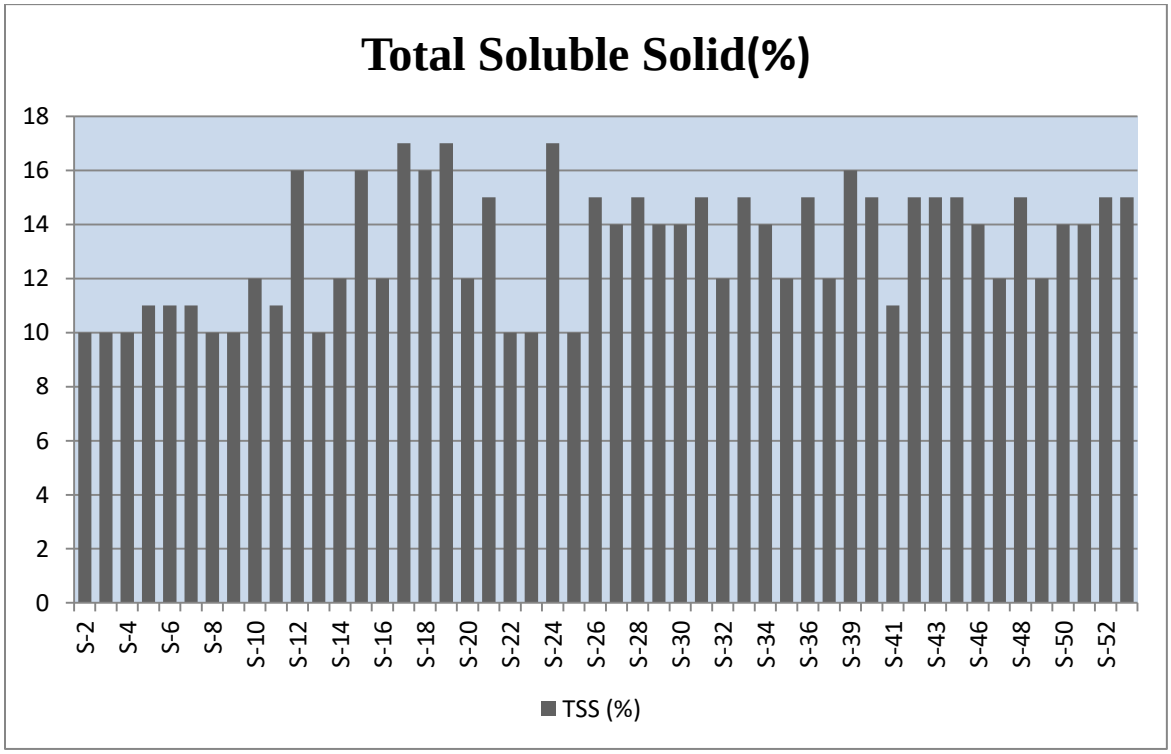


Figure 4. 8 Graphical presentation of the variation of TSS content with the different mango varieties.

pH Content

pH content is commonly the index of acidity parameter, which is an important internal parameter of any fruit. pH content depends on the maturity of the mango fruit and mature

fruits can be sorted and graded for the export purpose. Table 1 depicts the highest value of pH for mango is 4.79 and the lowest value for the same is 2.74. pH values less than 7 are defined as acidic and pH values more than 7 are defined as alkaline or basic. As, in this study it has been observed that different varieties of mango has a range from 2.74 to 4.79 of pH value, it can be concluded that mango is an acidic fruit. From the descriptive statistics the mean value of pH content has concluded as 3.975. Kalapahar cultivar from Wes Bengal and Imampasand cultivar of Maharashtra is having the highest and lowest percentage of pH respectively. The trend from this study has been shown that most of the variety contains the pH more than 4. The variation of the pH content for different varieties has been reasoned for different ripening period of concerned geographical area of current study [1]. A graphical data representation has been done in Figure 4.9 that is the variation of moisture content with different varieties.

Present research work expressed a range of different values of concerned nutritional parameters. This variation of the nutritional properties occurs generally due to some reasons. The reasons might be soil contamination; the quality of soil differs with the places. The storage temperature and ripening process are responsible for the enormous metabolic activities within the mango fruits, which can lead the chemical changes from variety to variety of different region [1, 16].

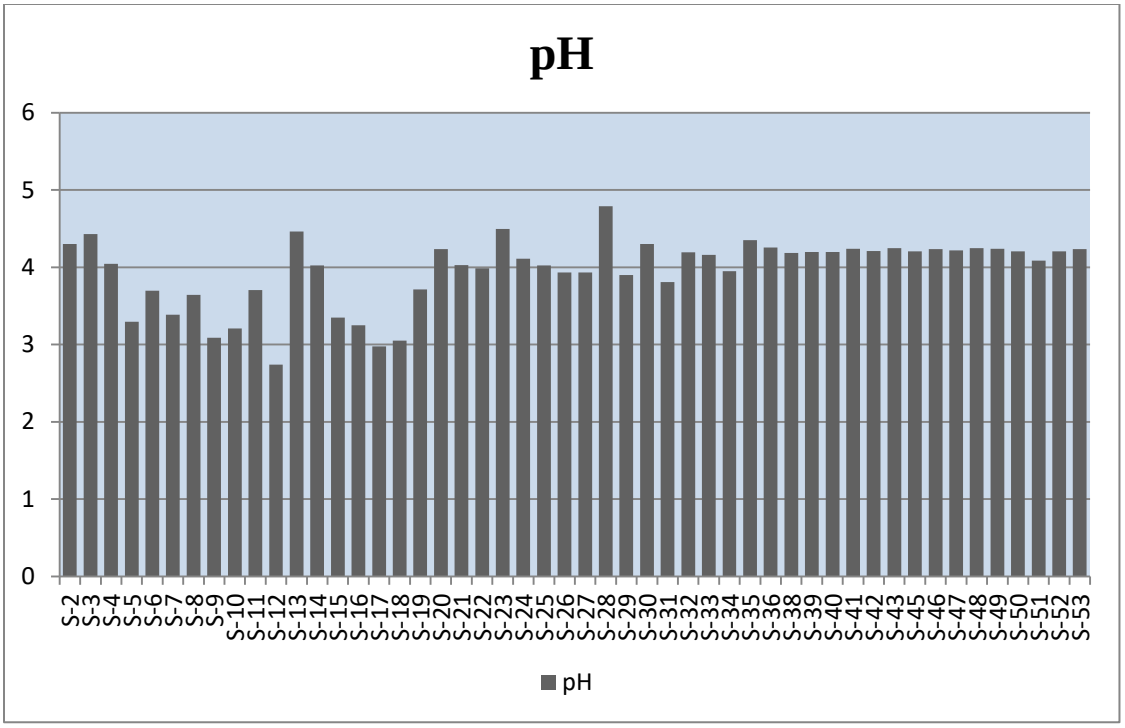


Figure 4. 9 Graphical presentation of the variation of pH content with the different mango varieties.

4.3.3. Determination of Rheological Properties

A descriptive statistics of the rheological parameter has been presented in Table 4.3 and it has been shown that the viscosity of the mango pulp varying from 8.8 Cst (Centistoke) to 257.8 Cst. The variation is very high and it is indicated that the Indian mango varieties have different flow properties. Some varieties have very thin and liquid kind pulp and many varieties have thick and slurry pulp.

Table 4. 3 Descriptive statistics of rheological parameters of 52 different samples of Indian mango using cup flow method

Rheological parameter	Mean	Standard Deviation	Maximum	Minimum	Coefficient of variance
Viscosity (Centi stoke)	89.183	60.215	257.8	8.8	0.675

4.3.4. Correlation between the Physical and Chemical Parameters

Correlation analysis among all the determined physical and chemical parameters of mango has been analyzed with Kearl Pearson correlation method or commonly known as covariance method and presented in Table 4.3 and Table 4.4.

This data has shown (Table 4.4) there is a strong correlation of the weight of intact fruit with size of the fruit, pulp weight and surface area of the fruit; pulp weight with the size of the intact fruit and surface area; size with surface area. The strong correlation has considered with the values of correlation coefficient greater than 1. The graphical presentation has been shown in Figure 4.10-4.15.

Intact fruit weight is positively correlated with pulp content, size of intact mango and surface area of intact mango with R^2 value of 0.977, 0.945 and 0.962 respectively. The pulp content is also positively correlated with the size of intact mango and surface area of the intact mango with R^2 value of 0.912 and 0.931 respectively. There is also a positive correlation between size and surface area of intact mango with R^2 value of 0.997. In terms of the percentage of variation in the dependent variables that is explained by the fitted sample regression equation, the Coefficient of Determination R-square quantifies the goodness-of-fit of the estimated Sample Regression Plane (SRP). Thus, the values of R square is 0.977, 0.945, 0.961, 0.912, 0.931 and 0.997 are simply means that about 97.7%, 94.5%, 96.1%, 91.2%, 93.1% and 99.7% of the variation in intact fruit weight is explained by the estimated SRP that uses pulp content, size and surface area, the variation in pulp weight is explained by the estimated SRP that uses size and surface area and the variation in size is explained by the estimated SRP that uses surface area respectively and R square value is significant at 1 % level.

Table 4. 4 Correlation coefficients between the physical parameter of mango

	Intact weight	seed weight	Peel weight	Pulp Weight	Size	Surface area	Sphericity
Intact weight	1.000						
seed weight	0.706	1.000					
Peel weight	0.610	0.486	1.000				
Pulp Weight	0.977	0.605	0.449	1.000			
Size	0.945	0.724	0.591	0.912	1.000		
Surface area	0.962	0.723	0.600	0.931	0.997	1.000	
Sphericity	0.310	0.095	0.096	0.342	0.171	0.197	1.000

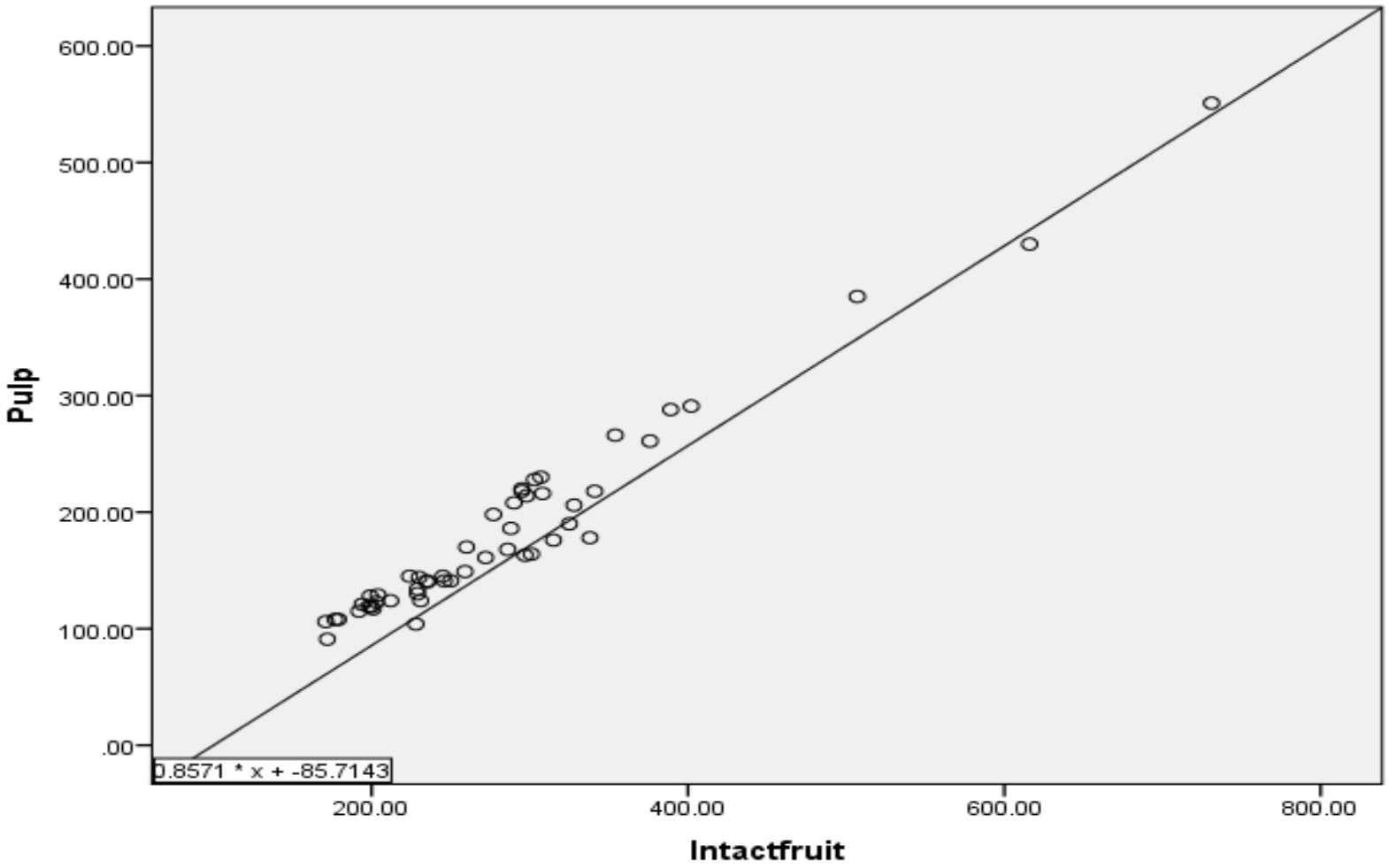


Figure 4. 10 Correlation between intact mango weight and pulp content

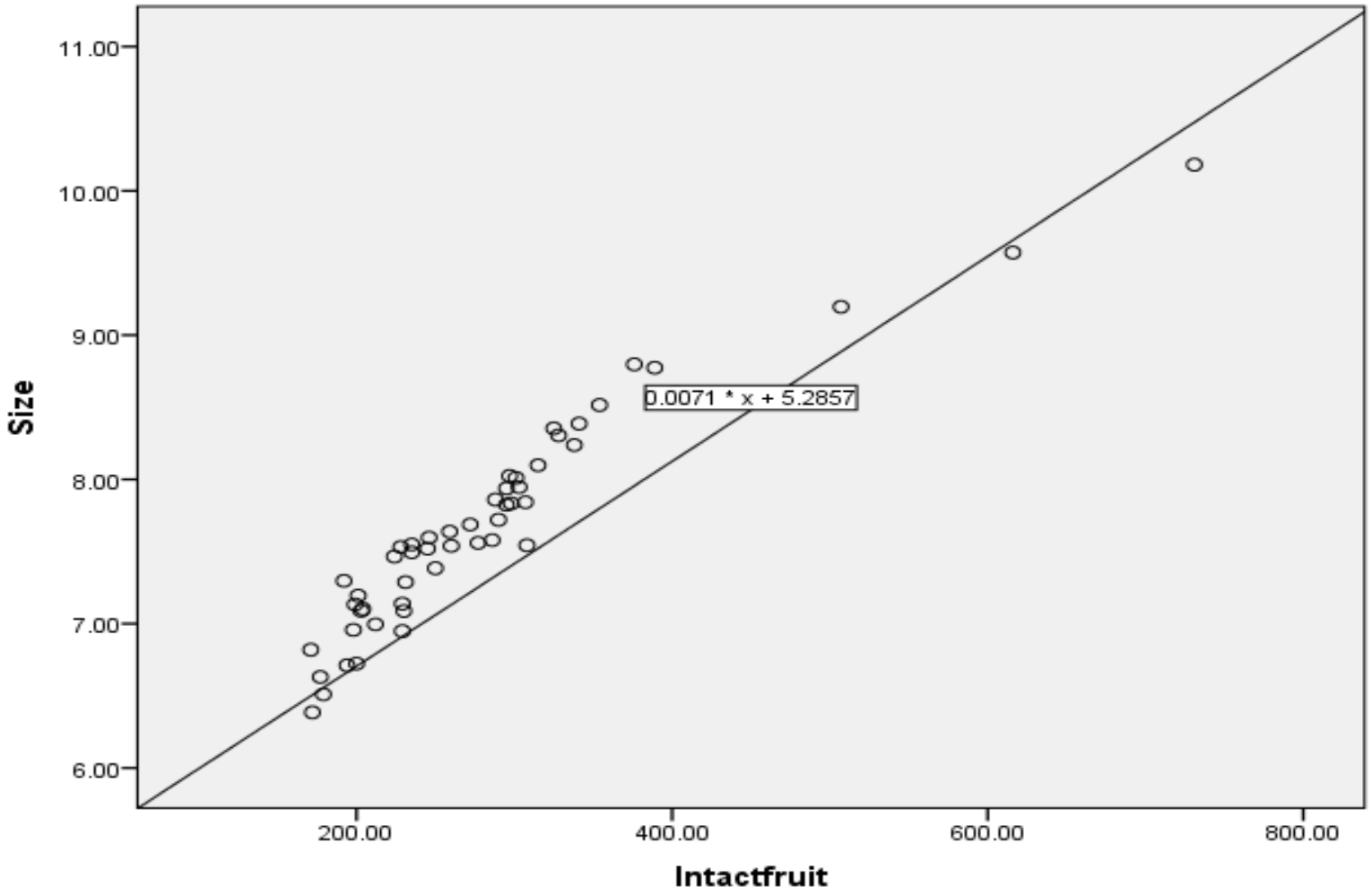


Figure 4. 11 Correlation between intact mango weight and size of the fruit

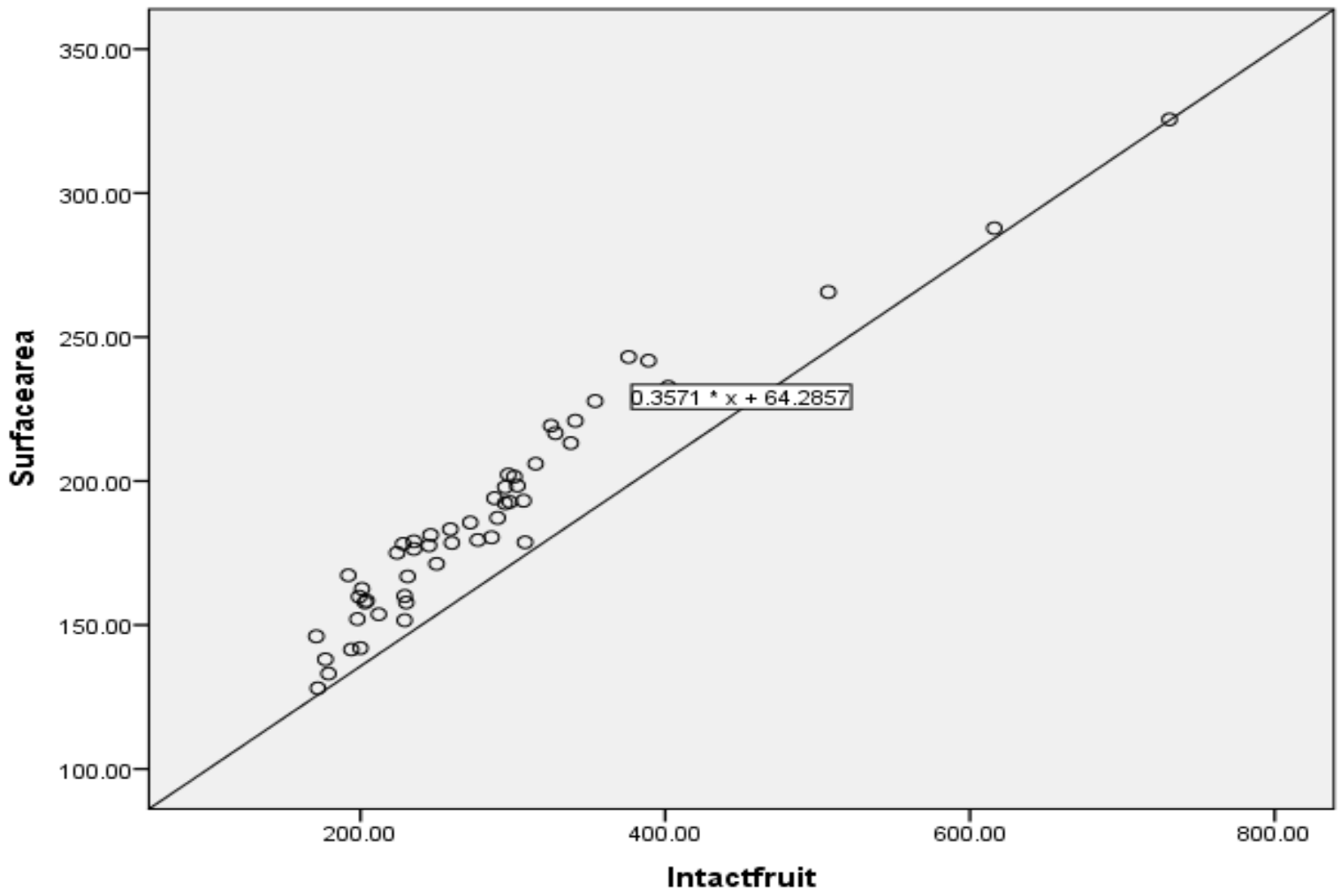


Figure 4. 12 Correlation between intact fruit weight and surface area of the fruit

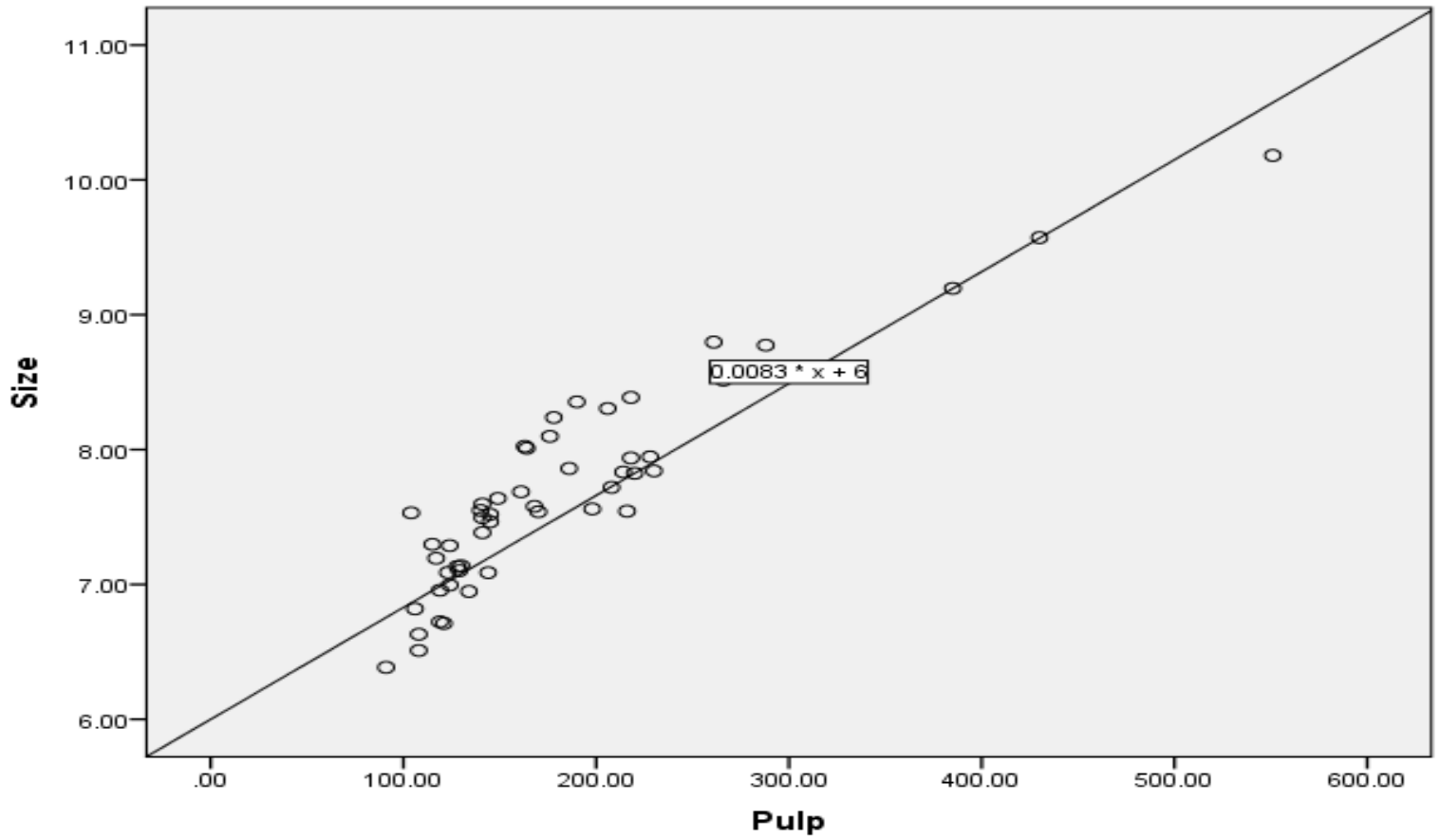


Figure 4. 13 Correlation between pulp content and size of the intact fruit

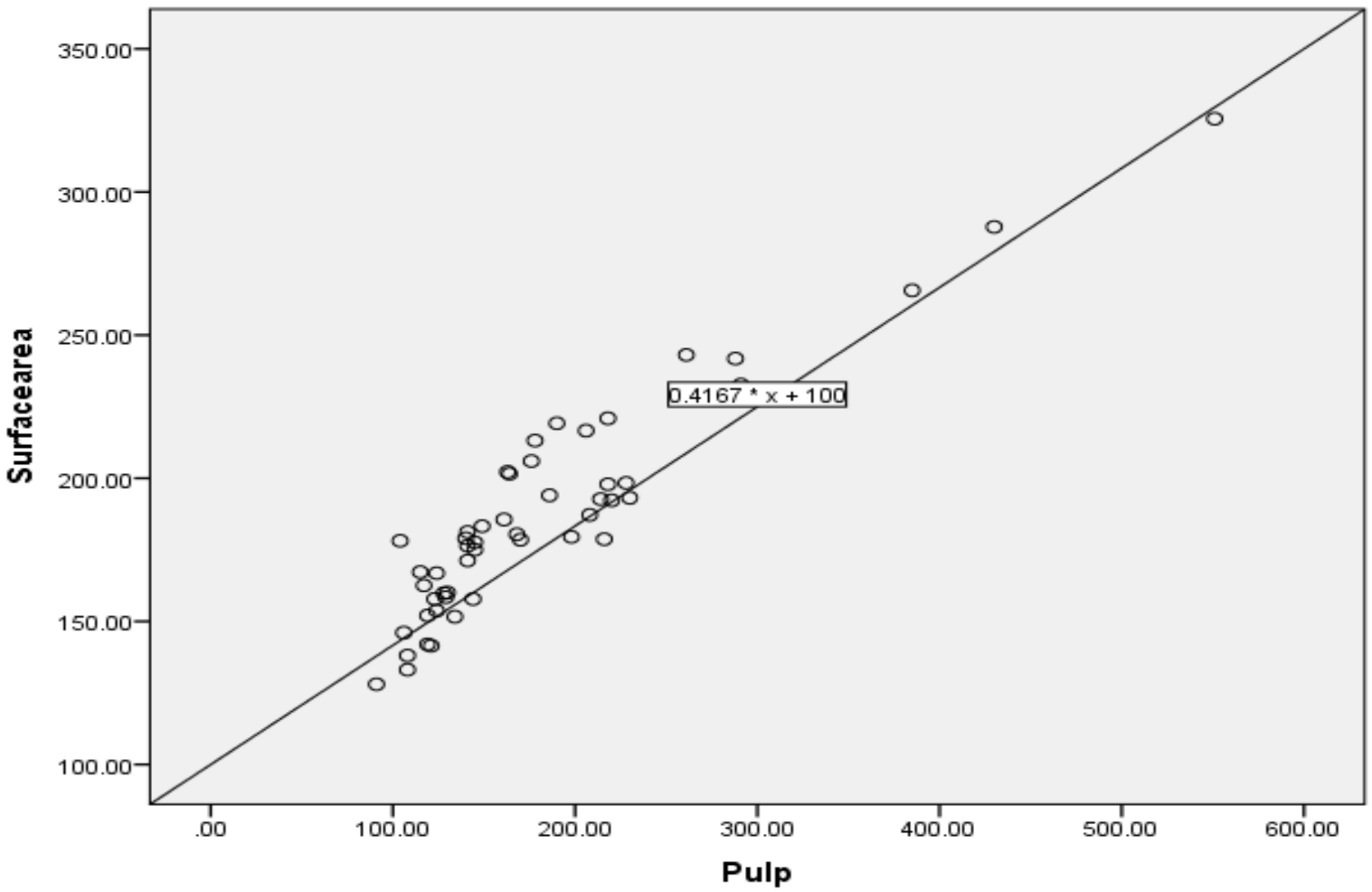


Figure 4. 14 Correlation between pulp content and surface area of the intact fruit

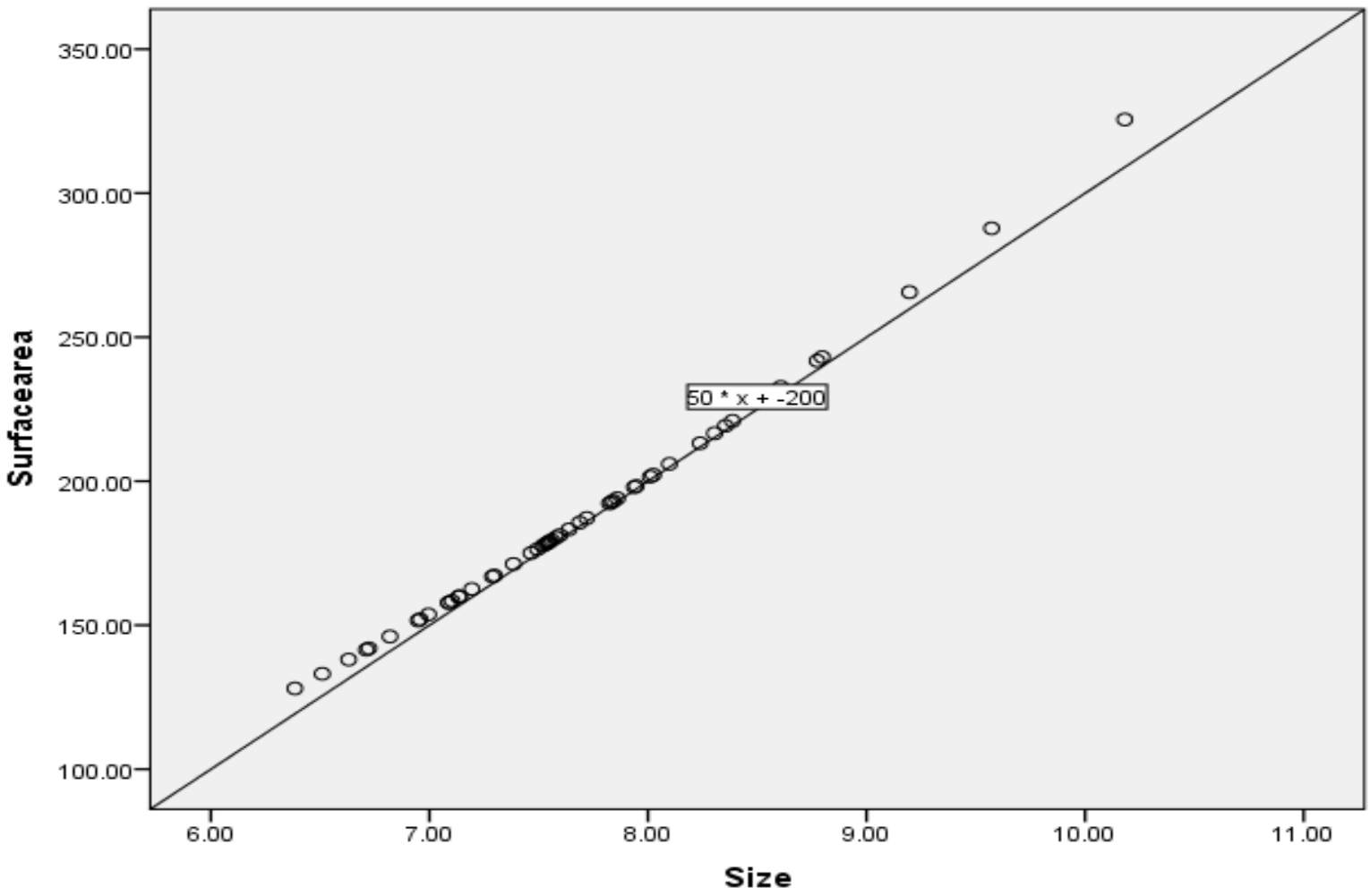


Figure 4. 15 Correlation between size and surface area of the intact fruit

The regression equation has been shown below,

Pulp content = 85.714 + 0.8571 fruit weight (26)

Size = 5.2857 + 0.0071 fruit weight (27)

Size = 6.000 + 0.0083 pulp content (28)

Surface area = 100 + 0.4167 pulp content (29)

Surface area = 200 + 50 size (30)

Surface area = 64.2857 + 0.3571 fruit weight (31)

The determination of correlation coefficient (r) has concluded that there is a strong correlation with moisture between carbohydrate and energy respectively and also between carbohydrate and energy. The correlation method was carried out with the all 51 samples. The correlation has been applied on the each nutritional parameters of mango pulp. The correlation coefficients between all the nutritional parameters have been presented (Table. 4.5). From this table it has been shown that there is strong correlation between moisture and carbohydrate, moisture and energy and carbohydrate and energy with r value greater than 0.9. The graphical presentation has been shown in Figure 4.16-4.18. Further regression analysis has been applied to know the equation between those parameters, which has been predicted their relationship mathematically. The regression equation has been shown below

Carbohydrate = - 0.9 Moisture + 84.9 (32)

Energy = -4.01 Moisture + 3.98 e^2 (33)

Energy = 4.15 Carbohydrate + 24.05 (34)

Energy is negatively and positively correlated with moisture and carbohydrate with R^2 value of 0.919 and 0.936 respectively and carbohydrate is negatively correlated with moisture with R^2 value of 0.861. The Coefficient of Determination R-square measures the goodness-of-fit of the estimated Sample Regression Plane (SRP) in terms of the proportion of the variation in the dependent variables explained by the fitted sample regression equation. Thus, the values of R square is 0.919, 0.936 and 0.861 are simply means that about 91.9%, 93.6% and 86.1% of the variation in moisture is explained by

the estimated SRP that uses energy, the variation in carbohydrate is explained by the estimated SRP that uses energy and the variation in moisture is explained by the estimated SRP that uses carbohydrate respectively and R square value is significant at 1 % level.

Table 4. 5 Correlation coefficients between the nutritional parameter of mango

	Moisture	Ash	Fat	Protein	Fiber	pH	Carbohydrate	Total Soluble Solid	Energy
Moisture	1.000								
Ash	-0.017	1.000							
Fat	-0.125	-0.149	1.000						
Protein	-0.403	0.029	0.454	1.000					
Fiber	0.027	0.002	0.382	0.270	1.000				
pH	-0.153	-0.053	0.092	0.123	0.436	1.000			
Carbohydrate	-0.928	-0.278	0.024	0.160	-0.122	0.133	1.000		
TSS	-0.047	0.163	0.402	-0.048	0.360	-0.070	-0.012	1.000	
Energy	-0.959	-0.265	0.199	0.394	-0.026	0.157	0.967	0.013	1.000

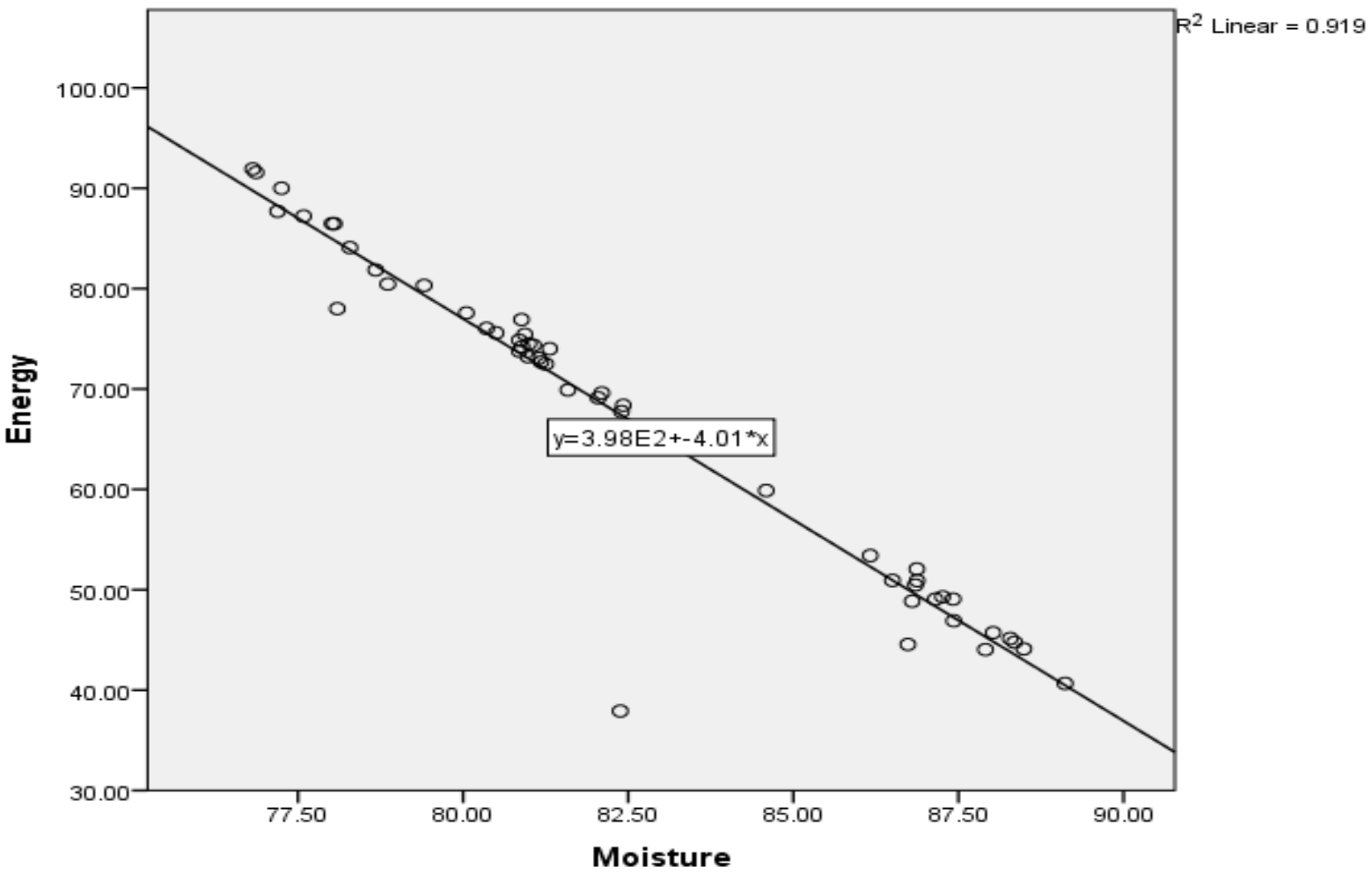


Figure 4. 16 Correlation between moisture and energy

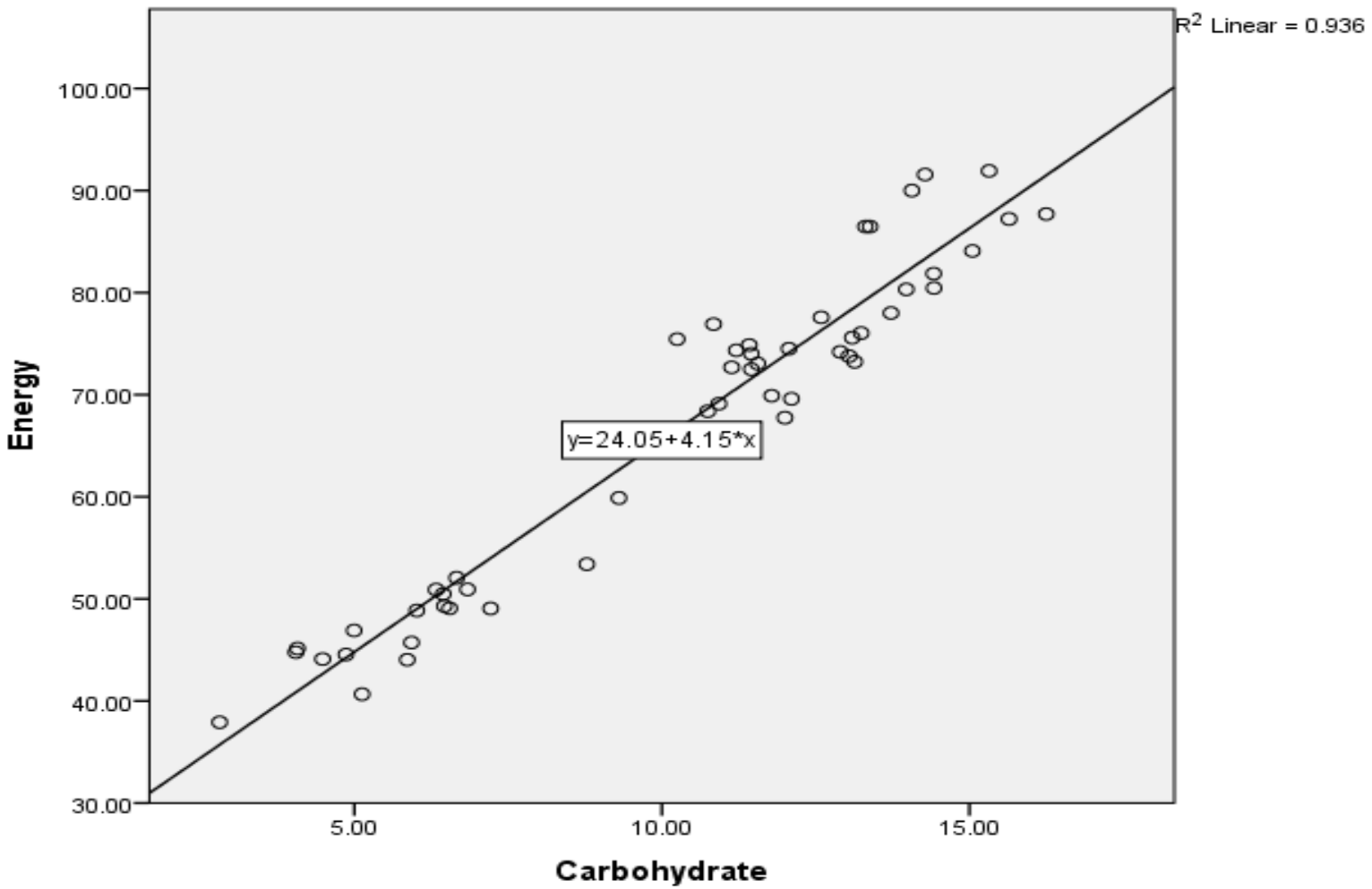


Figure 4. 17 Correlation between carbohydrate and energy

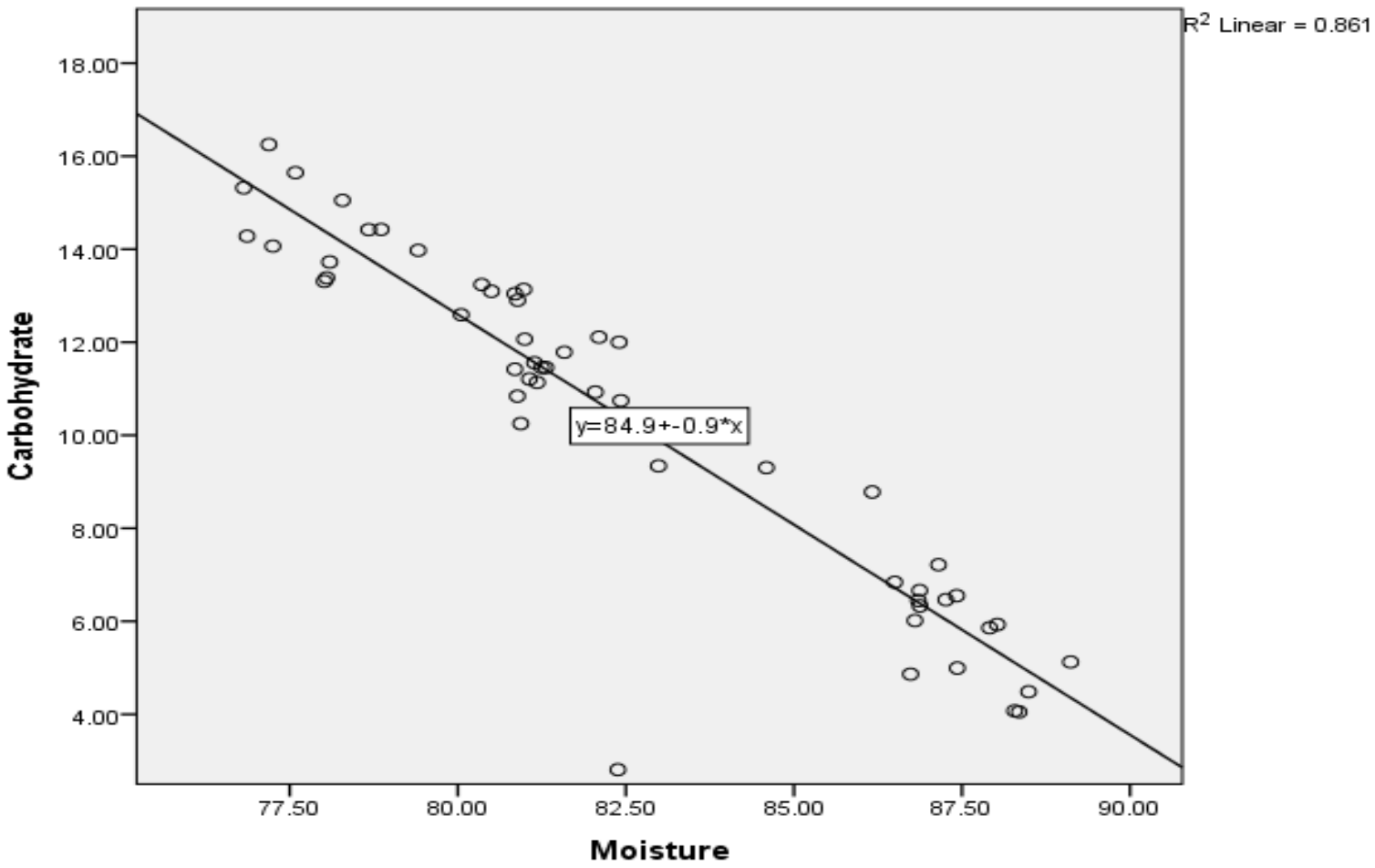


Figure 4. 18 Correlation between moisture and carbohydrate

4.4. Conclusion

The outcomes from this chapter have an efficient impact on the food science and nutritional studies. The physical, nutritional and rheological attributes of the enormous varieties of mangoes has been calculated with this study and the descriptive statistics has been explained. The descriptive statistics has concluded the range of all parameter with minimum and maximum values. Correlation between all those parameters was observed and it was noticed that energy, moisture and carbohydrate content and pulp content, weight, size and surface area of the mango fruit showed the best correlation. The best correlation for all the cases showed the values of correlation coefficient more than 0.9. Energy is negatively and positively correlated with moisture and carbohydrate with R^2 value of 0.919 and 0.936 respectively and carbohydrate is negatively correlated with moisture with R^2 value of 0.861. Intact fruit weight is positively correlated with pulp content, size of intact mango and surface area of intact mango with R^2 value of 0.977, 0.945 and 0.962 respectively. The pulp content is also positively correlated with the size of intact mango and surface area of the intact mango with R^2 value of 0.912 and 0.931 respectively. There is also a positive correlation between size and surface area of intact mango with R^2 value of 0.997. Also a statistical relation between energy, moisture, carbohydrate and pulp content, weight, size and surface area of intact fruit has been developed for the mango pulp. Those predicted equations have been proved as time effective mathematical equations with which the two parameters can be determine with one parameter.

References

1. S. Akhtar, M. Riaz, A. Ahmad and A. Nisar, *Pak. J. Bot*, **42**, 853-862 (2010).
2. V. R. Lebaka, Y. J. Wee, W. Ye and M. Korivi, *Int. J. Environ. Res. Public Health*, **18**, 741 (2021).
3. T. M. M. Malundo, R. L. Shewfelt, G. O. Ware and E. A. Baldwin, *J. Am. Soc. Hortic. Sci.*, **126**, 115-121 (2001).
4. A. Swaroop, M. Bagchi, H. Moriyama and D. Bagchi, *Jpn J Med*, **1**, 149-154 (2018).
5. FAO, 2019. Tropical fruits compendium. http://www.fao.org/docrep/meeting/022/am481t.pdf.
6. G. Girma, G. Garo and S. Fetena, *Int. J. Nutr. Food Sci*., **5**, 395 (2016).
7. P. Mamiro, L. Fweja, B. Chove, J. Kinabo, V. George and K. Mtebe, *Afr. J. Biotechnol*., **6**, 2477-2483 (2007).
8. U. P. Singh, D. P. Singh, M. Singh, S. Maurya, J. S. Srivastava, R. B. Singh and S. P. Singh, *Int. J. Food Sci. Nutr.*, **55**, 163-169 (2004).
9. T. Yantarasri, S. Ben-Yehoshua, V. Rodov, W. Kumpuan, J. Uthaibutra, J. Sornsrivichai *Postharvest Physiol. Fruit*., **398**, 81-92 (1994).
10. S. N. Jha, A. R. Kingsly, S. Chopra, *J. Food Eng*., **72**, 73-76 (2006).
11. T. Osadare, O. Koyenikan, F. Akinola, *Asian Food Sci. J.*, **10**, 1-8 (2019).
12. J. C. Ehiem, K. J. Simonyan, *Int. Agrophys*., **26**, 95-98 (2012).
13. I. S. Ashoush, M. G. Gadallah, *World J. Dairy Food Sci*., **6**, 35-42 (2011).
14. H. K. Sreenath, A. M. Nanjundaswamy and K. R. Sreekantiah, *J. Food Sci.*, **52**, 230-231 (1987).
15. AOAC. (2000). Official Methods of Analysis of the Association of Official Agricultural Chemists, Washington, DC, pp-125-139.
16. M. Shahnawz, S. A. Sheikh and S. G. Khaskheli, Afr. J. Biotechnol., **11**, 9825-9828 (2012).

Chapter-5: Classification of Mango Using PCA

Chapter-5

Classification of Mango (*Mangifera indica*) Using PCA

This chapter focuses on the geographical classification of mangoes on the basis of three obtained properties that is physical properties, chemical properties and the spectral properties. Spectral properties have obtained into two parts those are near infrared (NIR) spectra and Fourier Transform infrared (FTIR) spectra. The classification has been performed with the Principal Component Analysis (PCA).

5.1. Introduction

Foothills of Himalayan range of the southern Asia that is India, Burma, Indonesia is basically known as the origin of mango fruits [1]. Mango is recognized as a "king of fruit" of its enormous attributes. Juicy flavor, excellent aroma, versatile shades of skin color, sweetness and the presence of essential nutrients has made "King of fruits" as the best Indigenous tropical fruits among all. India has the sufficient varities to produce the world market of need. More than hundred varities have been cultivated within the Indian subcontinent [2]. Few of them are commercially used for the export and import purpose and for industrial use as well. Although mango has different varieties, all of them belong from same family Anacardiaceae [3]. The largest volume of the mango that is near about 64% has been produced by India. Also there are many food products which have been processed from mango pulp like jam, jelly, puree and pickle has been exported to all over the world [2, 4]. Mango is cultivated and domesticated from many decades within the Indian tropical region. The history of cultivation of agricultural crops and the civilization is more over associated with the mango cultivation from ancient time. Mango has been acknowledged with different cultural, religious and social economical values [1]. Maturity and quality of the mango fruit can be assessed with surface firmness, glossy skin texture and juicy flavor by maximum consumers. But, most of the time this kind of

assessment with only eye vision has proved as a mislead [5]. For Indian market mango has a great demand and also it has a worldwide export demand. The need to study of physical attributes of the mango fruits are important to providing valuable engineering data required for designing and to develop the machines, structures and equipment for transporting, which includes handling, transporting , processing and storage of mango fruits. Grading, sorting, packaging and cleaning of the mango fruits before export to various places from the different farms the two physical attributes i.e. shape and size strongly considerable [6, 7]. Other important physical properties like specific gravity and density is relevant to calculate the terminal velocity, thermal diffusivity in heat transfer and Reynolds number for pneumatic and hydraulic handling of products. Storage, transportation and separation system can be categorized with mass of the intact fruit and the bulk density and porosity. Almost all the physical parameters have its own mathematical expression for determination [8]. Shape is an important key parameter of mango fruits for the export purpose. Two components that is longitudinal and lateral cross section of the fruit consider as main components for shape determination [8, 9].

Infrared spectroscopy is a nondestructive approach to qualitative and quantitative analysis of any material. In the field of food science and food technology this method has achieved an important role to solve the several research problems [10]. Near Infrared spectroscopy is classified as a non-invasive, nondestructive, robust and fast method. It takes few seconds to analyze the samples and determine the presence of infrared active molecules. The main benefit of the non-destructive study is the minimal use of chemicals only light matter interaction is there and after NIR spectroscopic technique the samples can be classified with all the attributes in an easy manner [11]. Also this method is very efficient for handing and marketing process for the mango fruits. Grading and sorting is also an important key for exportation and can be determine easily rapidly determine with few scans. This method has proven as efficient tool for the qualitative and quantitative analysis of many fruits [12].

Infrared spectroscopy has commonly used for identification of organic compounds. Several functional groups like C-H, O-H, and N-H can be identified with the IR

spectroscopy and this has been proved as an attractive method for the several organic systems because of the user friendly sampling method [13].

Classification of the data is important to know the trend of any statistical data, and that classification can be done with PCA. This analytical method is the oldest multivariate analysis approach and was first invented by Jordan in 1874. Then in 1933 Hotelling and Wold have initialized the modern version of PCA with the addition of NIPLAS algorithm. This method basically has extracted the tabular data and made it representable with the new orthogonal variable that is principle component. Principle components are highly inter-correlated and dependent in nature. Pearson was elaborated the PCA with his own words that is “lines and planes of closest fit to systems of points in space”. Principle component analysis is also known as a dimension reduction method and it can reduce the large number of variables with small number of variables maintain the same information [14]. The principle component analysis is a mathematical tool which can apply to transform a number of correlated variables to uncorrelated variables. Those variables are known as principle components. This method is similar to the multivariate factor analysis and generally it applies on the square symmetric matrix [15]. A linear combination of observed variables is expressed as principal components. The main purpose of this analysis is to collect the correct information and ignore the outliers [16].

5.2. Materials and Method

5.2.1. Physical and Chemical Data Collection

Physical data of the intact mango has been determined with Vernier caliper and some mathematical formulations. Chemical data has been collected by performing AOAC standard methods. Details of these methods have been discussed in chapter-3.

5.2.2. Spectral Data Acquisition

After the physical property analysis the samples has been taken into the CSIR-CSIO Chandigarh for NIR collection and University Grant Commission Department of Atomic Energy Consortium for Scientific Research Centre (UGC DAE CSR) Kolkata for FTIR collection. The NIR and FTIR spectra data has been collected with NIR DOSS2500 instrument with spectral range of 700-2500 nm and with Perkin Elmer Fourier transform

infrared spectrometer with the KBr optical beam splitter and with 8300 to 350 cm^{-1} spectral range. Details of the NIR and FTIR spectrometer have been discussed in chapter-3.

5.2.3. Classification of Data

Principal Component Analysis (PCA) was employed using Singular value decomposition algorithm to classify the mango samples on basis of origin. Principal Component Analysis has been carried with the Clustvis 2.0 (https://biit.cs.ut.ee/clustvis/, accessed by 20 June, 2022) web tool for visualizing clustering and classifying of multivariate data.

5.3. Results and Discussions

5.3.1. Spectral Data

The original spectra of the raw NIR data of total 51 different mango samples was taken from the wavelength range of 700 to 2500nm (Figure 5.1). The spectra showed picks at different wavelengths indicating the presence of different organic molecule in the mango pulp sample (Figure 5.2).

The original spectra of the raw FTIR data of total 51 different mango samples was taken from the wavelength range of 700 to 2500nm (Figure 5.3). The spectra showed picks at different wavelengths indicating the presence of different organic molecule in the mango pulp sample (Figure 5.4).

Near infrared and Fourier Transform infrared spectroscopy both are the non-destructive and rapid techniques. But if the comparison occurs NIR is the best approach with some limitations. Both the radiations can vibrate the organic molecules and with the vibrational transitions Figure 5.1 and 5.2 has been recorded. With these two methods there no sample preparation has been done in the present research. NIR radiation can be much more penetrable into the sample as compare to FTIR radiation. As, FTIR can only penetrate at very shallow depth, it has shown the (Fig. 5.2) the fundamental organic bands. There are total four major peaks has been shown in figure 2. At 3394 cm^{-1}, a broad band has shown, which indicates presence of huge amount of water in the mango pulp. Generally at that peak many organic molecules can be vibrated that is O-H stretching, N-H stretching, B-O-H stretching and Si-OH stretching. At 2916 cm^{-1} aldehyde C-H stretching can be assigned. At 2137 cm^{-1} Azide N=N stretching, C=C stretching can be vibrated and

finally primary amide NH_2 bending band and C=C stretching band can be occurred at 1637 cm^{-1}. As NIR spectra can radiate into the higher depth into the sample as compare to FTIR radiations, it has shown (Figure 5.1) the harmonic and combination band of fundamental bands. But in NIR region the bands are not that sharp due to the higher frequencies. In Figure 5.1, second harmonic of O-H bond due to the presence of water in the sample has occurred at 977 nm. Second harmonic of C-H stretching band has occurred at 1193 nm due to the present of lipid in the mango pulp. First over tone of N-H stretching has been assigned at 1457 nm due to the presence of protein. First harmonic of C-H and combination of O-H band has been assigned at 178 nm and 1927 nm respectively due to the presence of cellulose and starch in mango pulp [17].

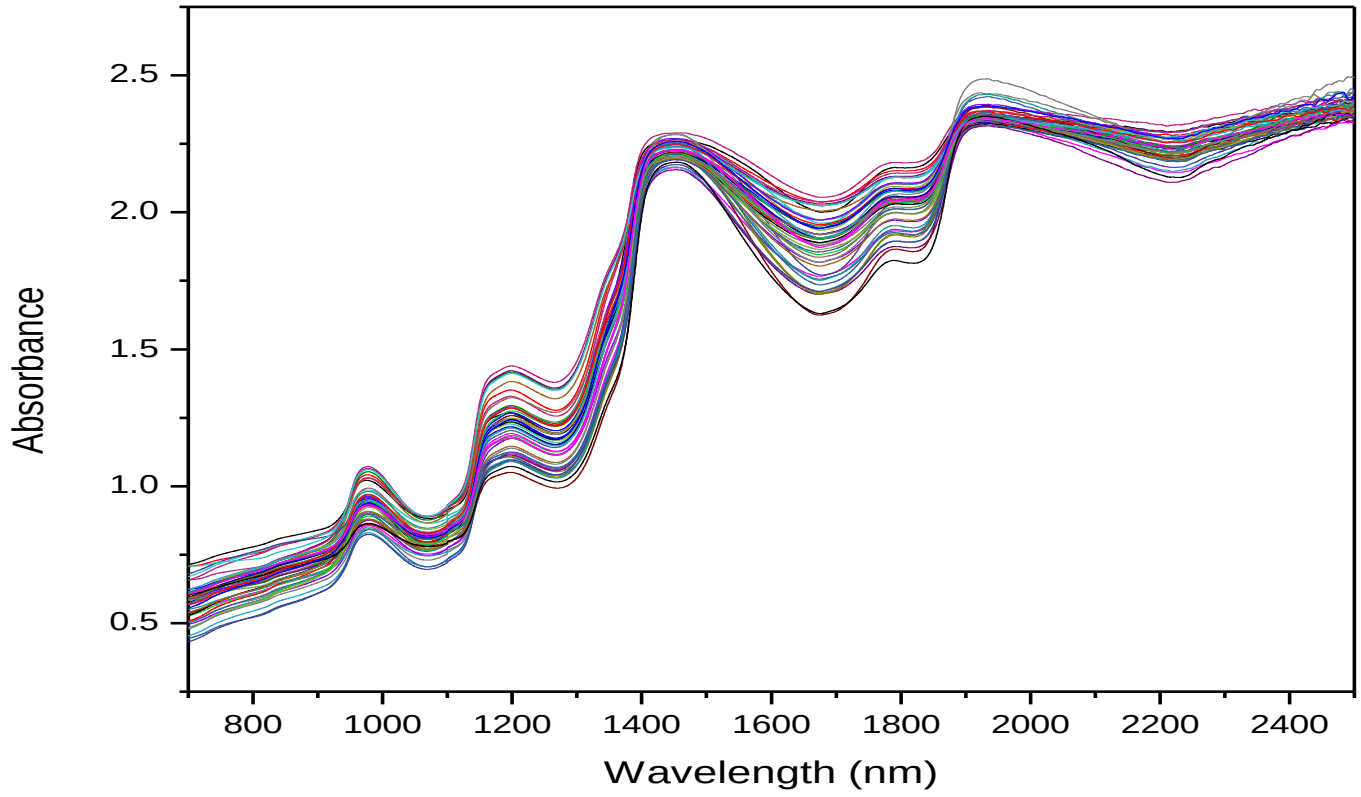


Figure 5. 1 NIR spectra of 51 mango samples

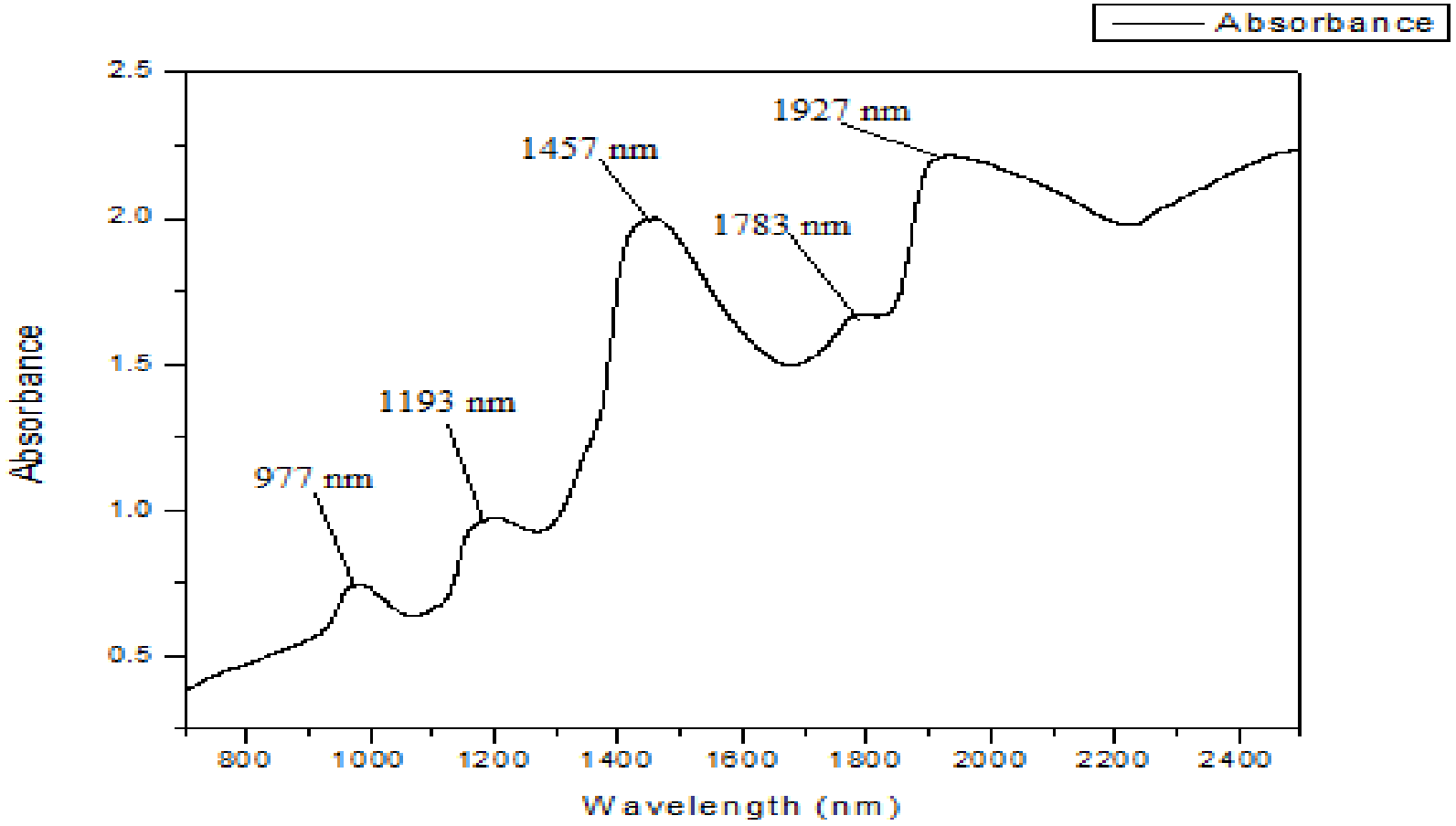


Figure 5. 2 Presence of the molecules in the mango pulp

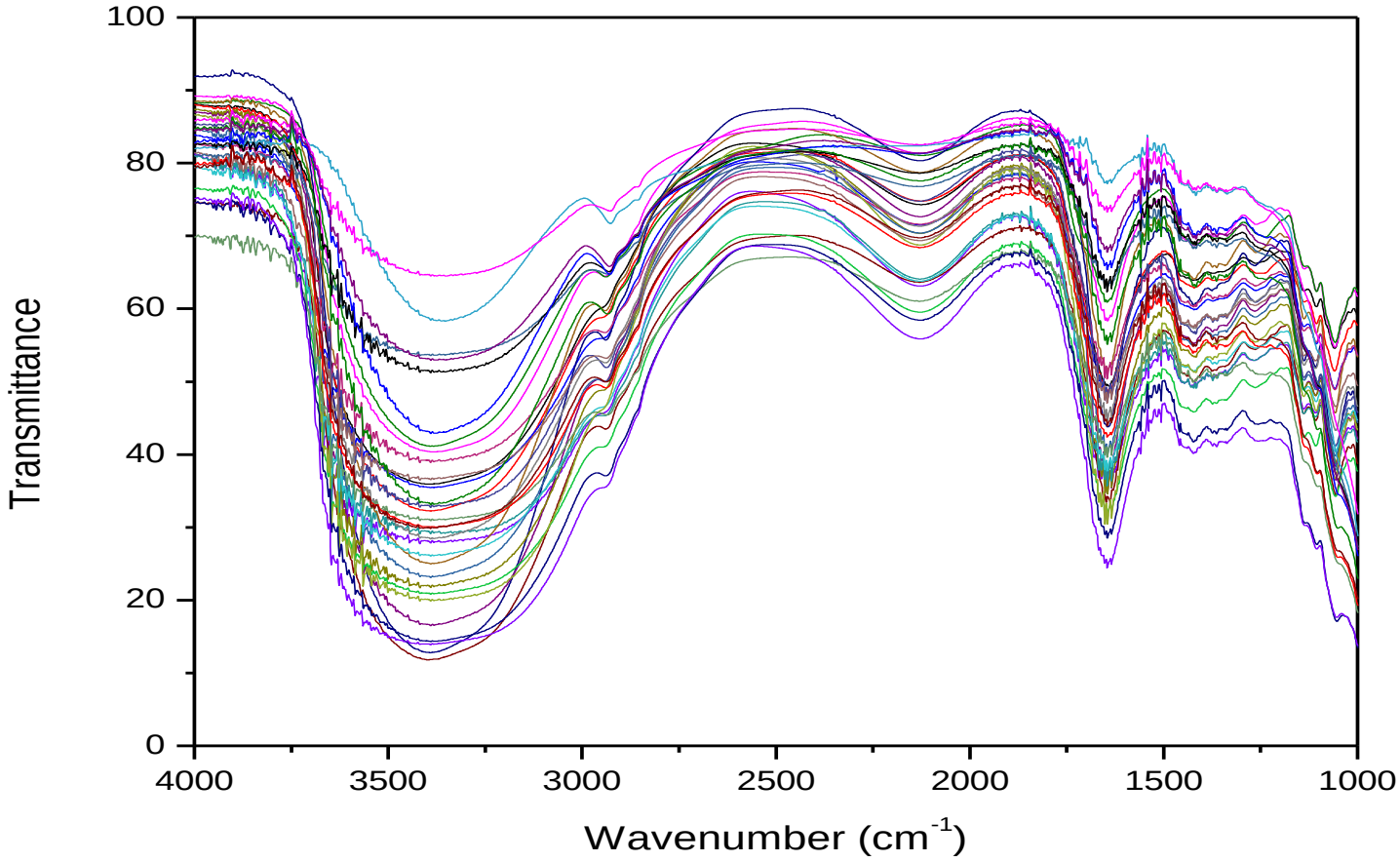


Figure 5. 3 FTIR spectra of 51 samples

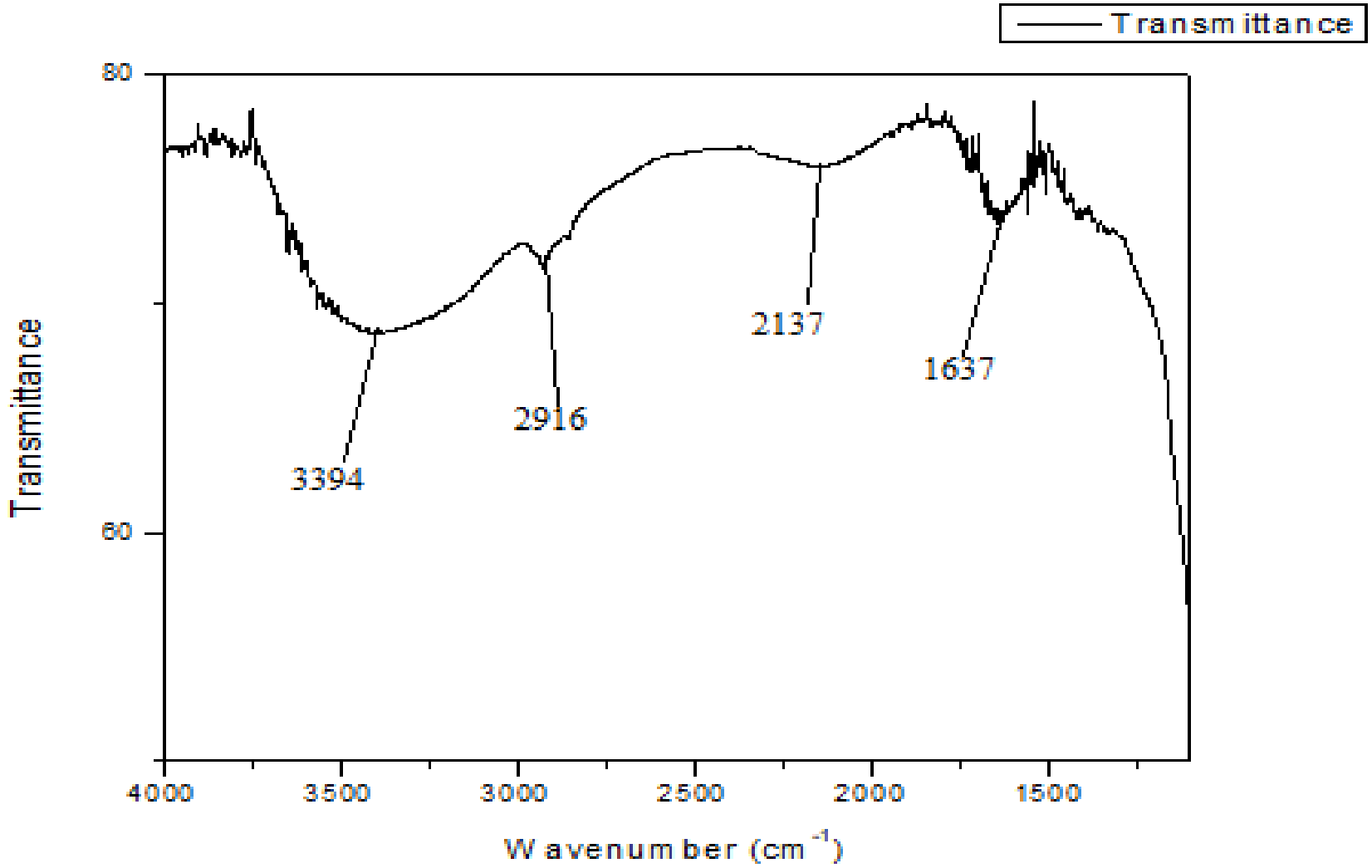


Figure 5. 4 Presence of the molecules in the mango pulp

5.3.2. Classification of Mango

Clustvis web tool application has been used to perform this statistical analysis. The chemical data has been inserted into the Clustvis in form of square symmetric matrix. Clustvis web tool application has been used to perform this statistical analysis. The chemical data has been inserted into the Clustvis in form of square symmetric matrix. A numeric data extracted as an input variable and performed PCA to reduce the factor. The PCA plots have been shown the scattered data between the principal components with the largest variance. There are three basic methods to calculate the variance of principal components that singular value decomposition (SVD) with imputation, non-linear iterative partial least square (NIPLAS) PCA and probabilistic PCA. The classification in present study has been determined with SVD with imputation method. PCA has been applied to experimental data to classify the data on region basis. A numeric data extracted as an input variable and performed principle component analysis (PCA) to reduce the factor. The PCA plots have been shown the scattered data between the principal components with the largest variance. There are three basic methods to calculate the variance of principal components that singular value decomposition (SVD) with imputation, non-linear iterative partial least square (NIPLAS) PCA and probabilistic PCA. The classification in present study has been determined with SVD with imputation method. PCA has been applied to experimental data to classify the data on region basis.

5.3.2.1. Classification on the Basis of Physical Parameters of Mango:

To classify the samples on the basis of physical properties, unit variance scaling was applied to rows; SVD with imputation was used to calculate principal components. X and Y axis showed principal component 1 and principal component 2 representing 68.7and 31.2% of the total variance for all 51 data points (all five states; Bihar, Maharashtra, Uttar Pradesh, Punjab and West Bengal: Figure 5.5), 81.6and 18.3% for 37 data points (three states; Bihar, Uttar Pradesh and West Bengal: Figure 5.6) and 70.2and 29.8% for 32 data points (only two states; Uttar Pradesh and West Bengal Figure 5.7), respectively.

The projected PCA in Figure 5.5 to Figure 5.7 has shown the fingerprint of the assessment of physical parameter classification. In Figure 5.5 and Figure 5.6 the

discriminant classification has not clearly shown, but in Figure 5.7, the classification of mango sample has been classified but not with all data points. So, it can be clearly concluded from these figures that the physical parameter of the varities of Indian mangoes are almost similar and distinct grouping for classification on the basis of physical properties is not giving any fruitful results.

5.3.2.2. Classification on the Basis of Chemical Parameters of Mango:

Similar classification has been done of the samples on the basis of nutritional properties, unit variance scaling was applied to rows; SVD with imputation was used to calculate principal components. X and Y axis showed principal component 1 and principal component 2 explaining 35.2 and 22% of the total variance for all 51 data points (all five states; Bihar, Maharashtra, Uttar Pradesh, Punjab and West Bengal: Figure 5.8), 36.5 and 21.5% for 37 data points (three states; Bihar, Uttar Pradesh and West Bengal: Figure 5.9) and 36.5 and 21.1% for 32 data points (only two states; Uttar Pradesh and West Bengal: Figure 5.10), respectively.

The projected PCA in Figure 5.8 to Figure 5.10 has shown the fingerprint of the assessment of physical parameter classification. In Figure 5.8 and Figure 5.9 the discriminant classification has not clearly shown, but in Figure 5.10, the classification of mango sample has been classified but not with all data points. So, it can be clearly concluded from these figures that the physical parameter of the varieties of Indian mangoes are almost similar and distinct grouping for classification on the basis of physical properties is not giving any fruitful results.

5.3.2.3. Classification on the Basis of FTIR Data of Mango:

Similar classification has been done of the samples on the basis of FTIR spectral data, unit variance scaling was applied to rows; SVD with imputation was used to calculate principal components. X and Y axis showed principal component 1 and principal component 2 explaining 75.1and 17.5% of the total variance for all 51 data points (all five states; Bihar, Maharashtra, Uttar Pradesh, Punjab and West Bengal: Figure 5.11), 72and 19.8% for 37 data points (three states; Bihar, Uttar Pradesh and West Bengal: Figure 5.12) and 74.5and 17.2% for 32 data points (only two states; Uttar Pradesh and

West Bengal Figure 5.13), respectively.

5.3.2.4. Classification on the Basis of NIR Data of Mango:

Similar classification has been done of the samples on the basis of NIR spectral data, unit variance scaling was applied to rows; SVD with imputation was used to calculate principal components. X and Y axis showed principal component 1 and principal component 2 explaining 65.6 and 21.4% of the total variance for all 51 data points (all five states; Bihar, Maharashtra, Uttar Pradesh, Punjab and West Bengal: Figure 5.14), 69.7 and 16.2% for 37 data points (three states; Bihar, Uttar Pradesh and West Bengal: Figure 5.15) and 59.7 and 28.6% for 32 data points (only two states; Uttar Pradesh and West Bengal Figure 5.16), respectively. For all the cases prediction ellipses are such that with probability 0.95, a new observation from the same group will fall inside the ellipse.

All the graphical results have been depicted that the classification of all the samples of five states cannot give any fruitful conclusion. But by reducing the data points finally with 32 samples of two different regions that is Uttar Pradesh and West Bengal can be classified with all attributes. Figure 5.7, 5.10, 5.13 and 5.16 has shown the classification.

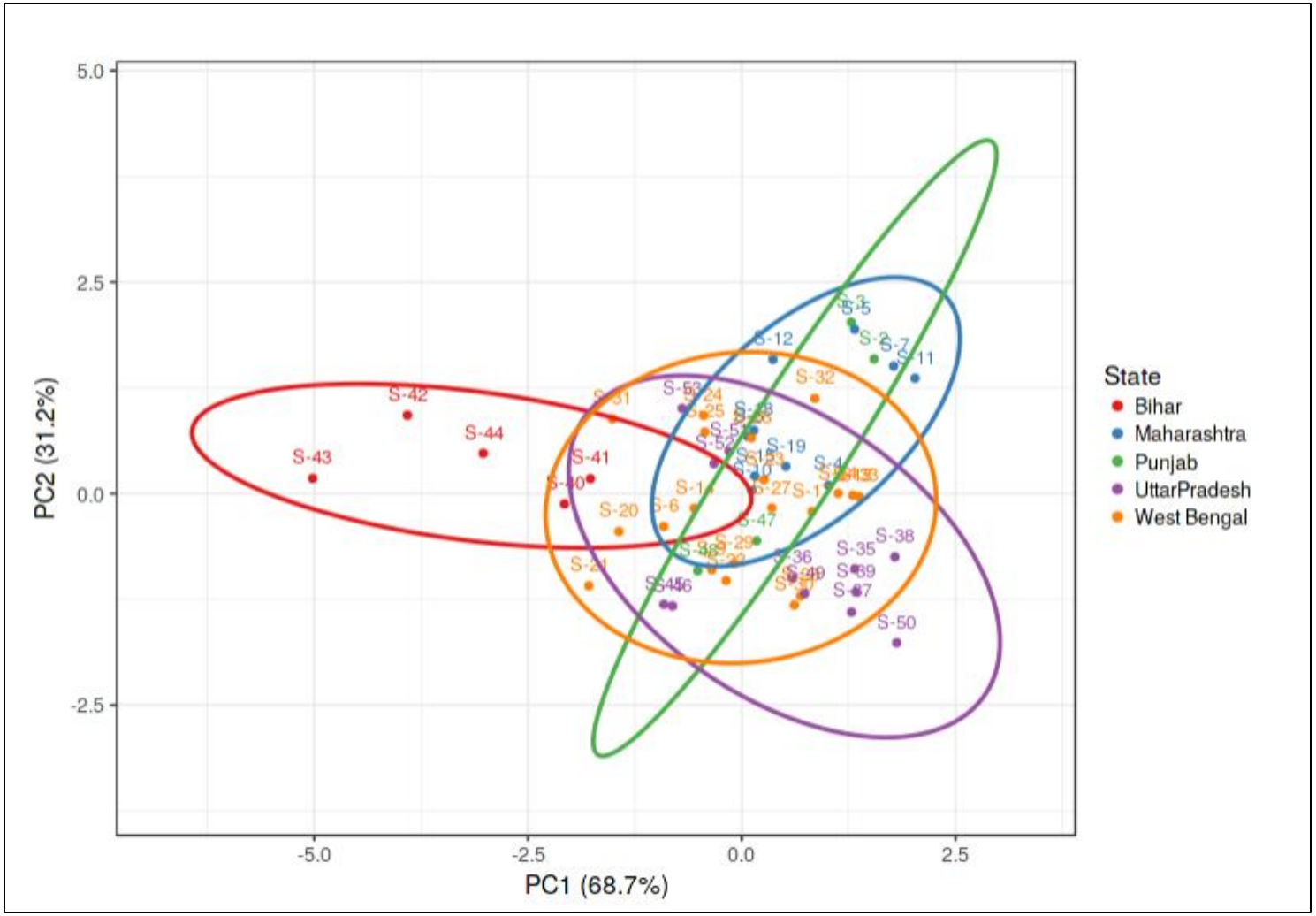


Figure 5. 5 Geographical (among five states) classification of mango on the basis of physical properties

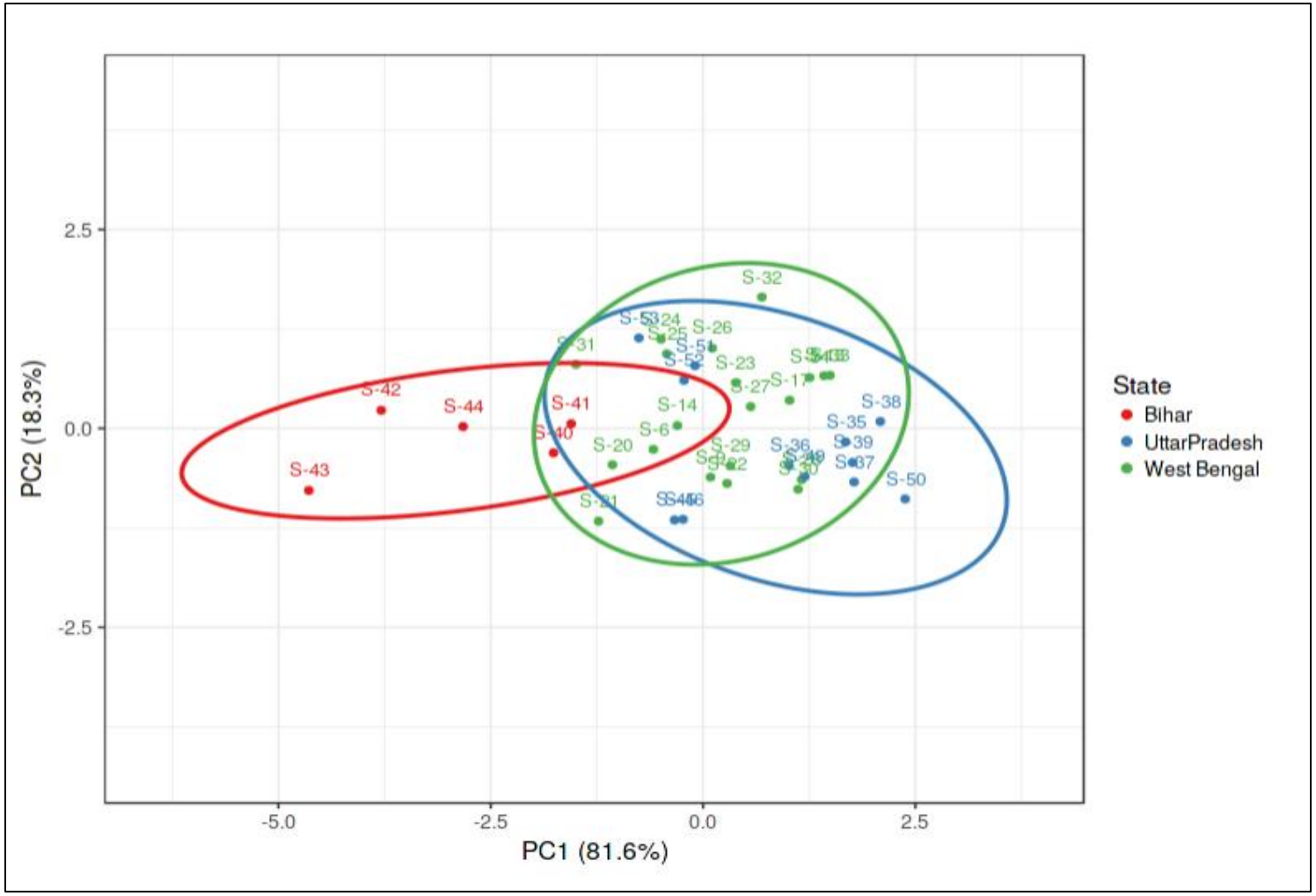


Figure 5. 6 Geographical (among three states) classification of mango on the basis of physical properties

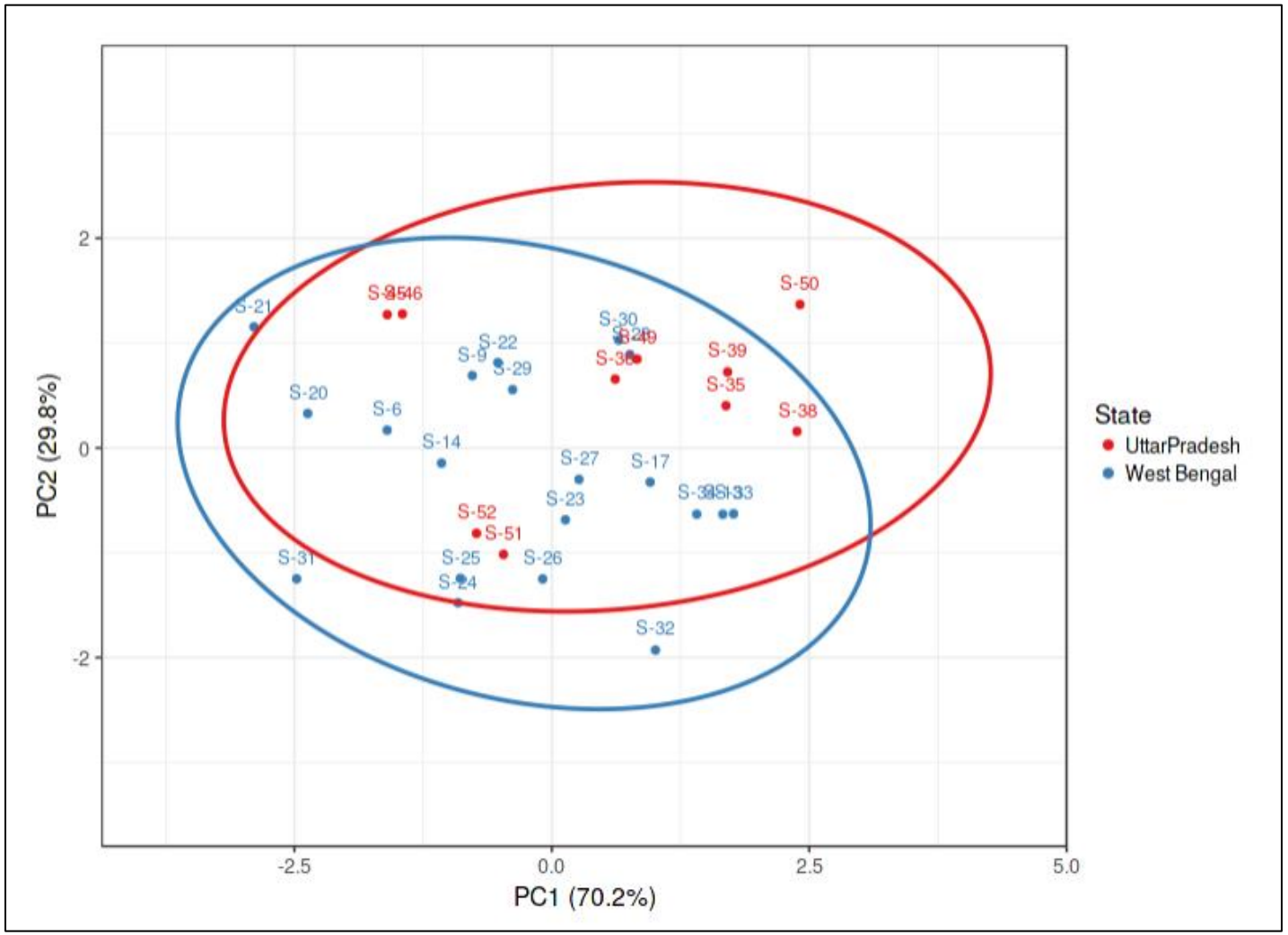


Figure 5. 7 Geographical (among two states) classification of mango on the basis of physical properties

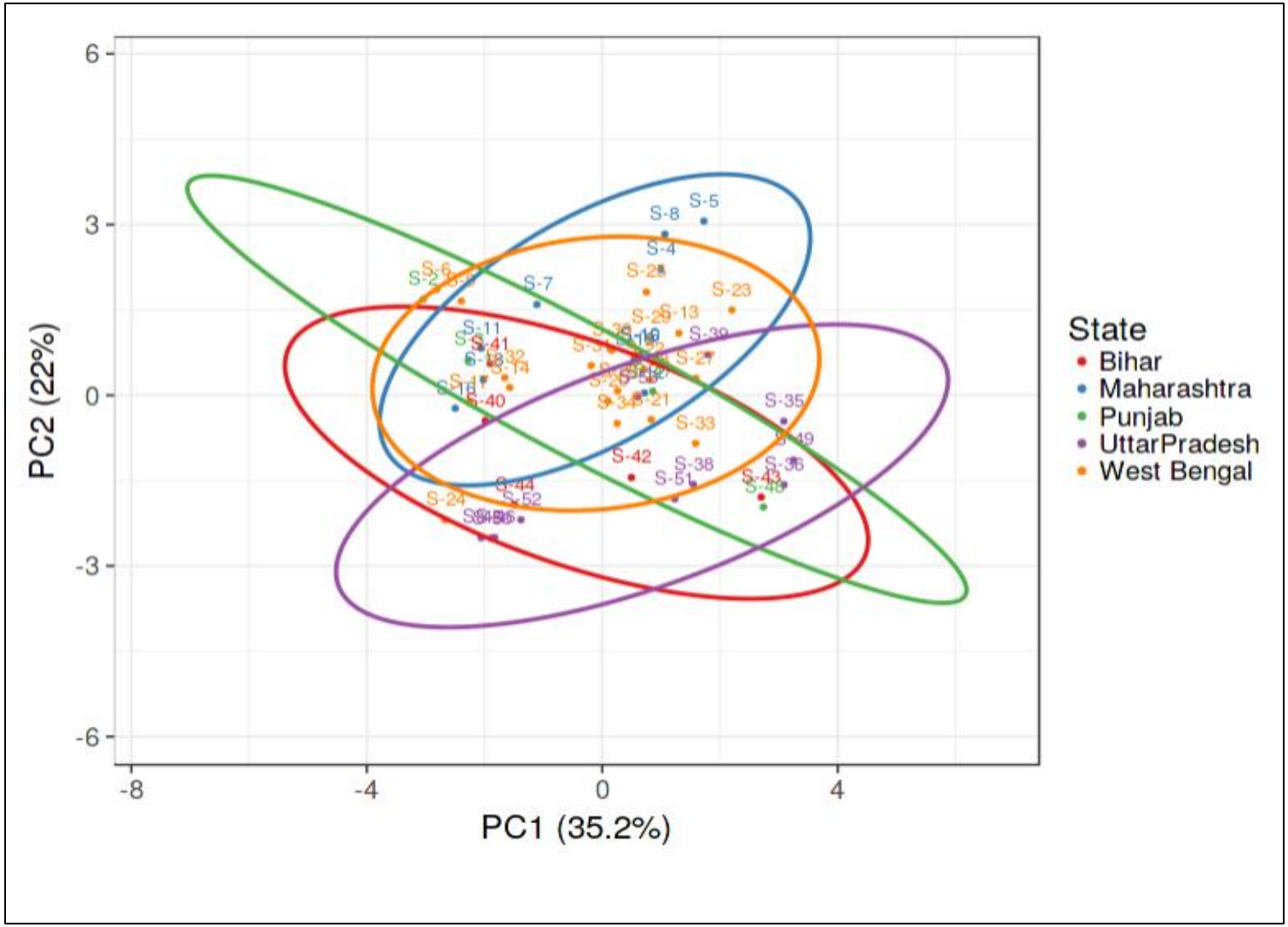


Figure 5. 8 Geographical (among five states) classification of mango on the basis of chemical attributes

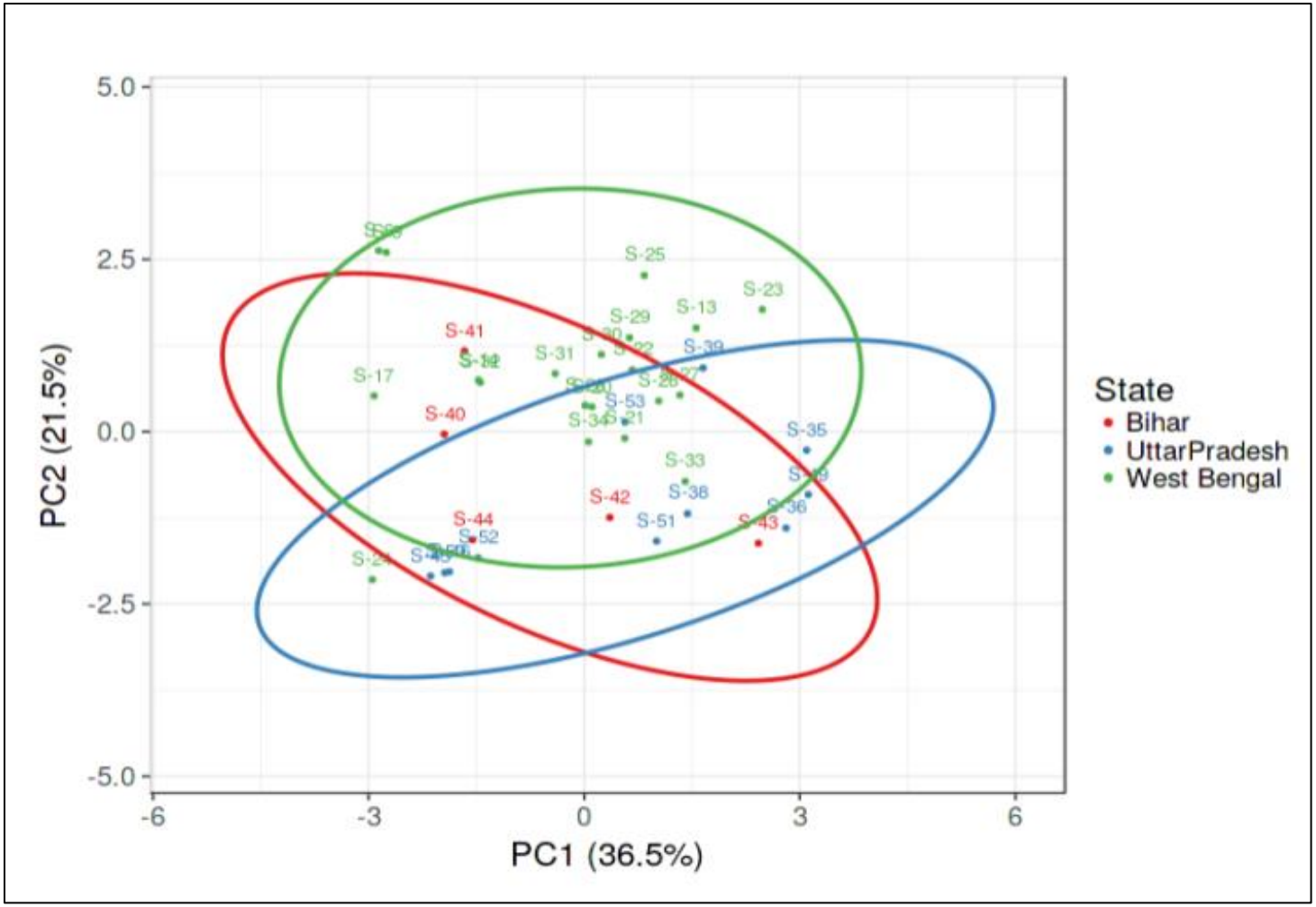


Figure 5. 9 Geographical (among three states) classification of mango on the basis of chemical properties

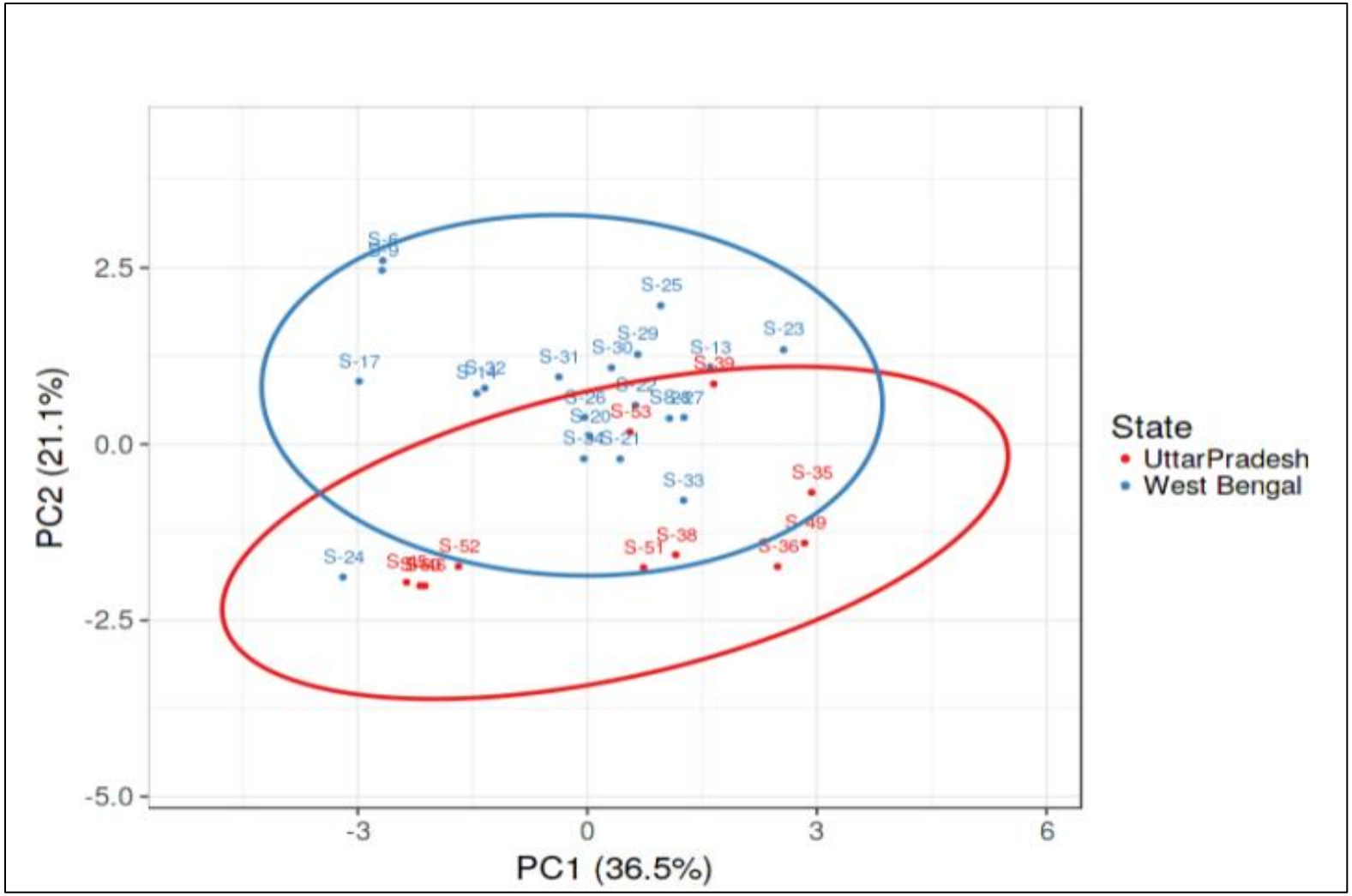


Figure 5. 10 Geographical (among two states) classification of mango on the basis of chemical properties

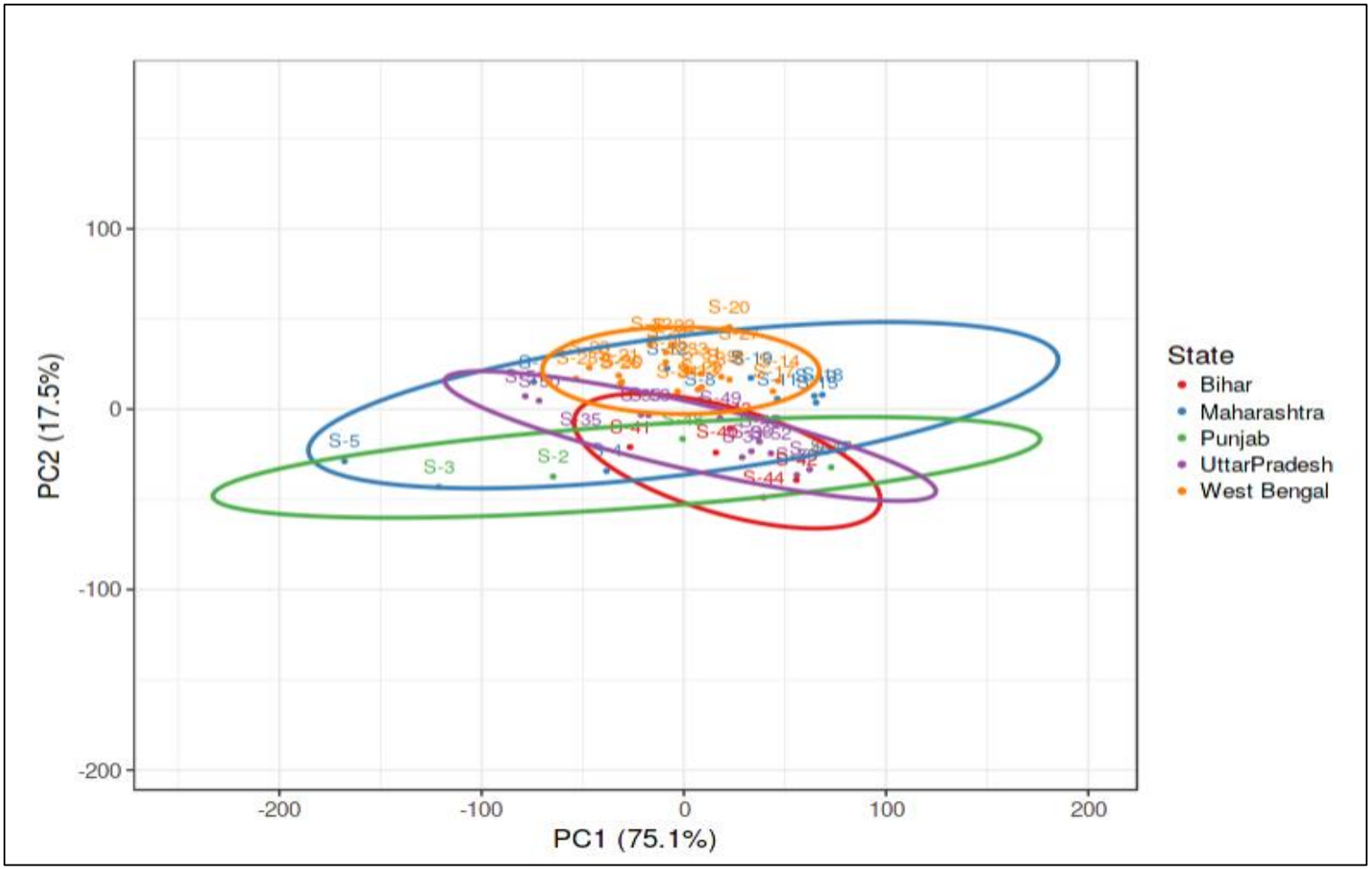


Figure 5. 11 Geographical (among five states) classification of mango on the basis of FTIR spectral data

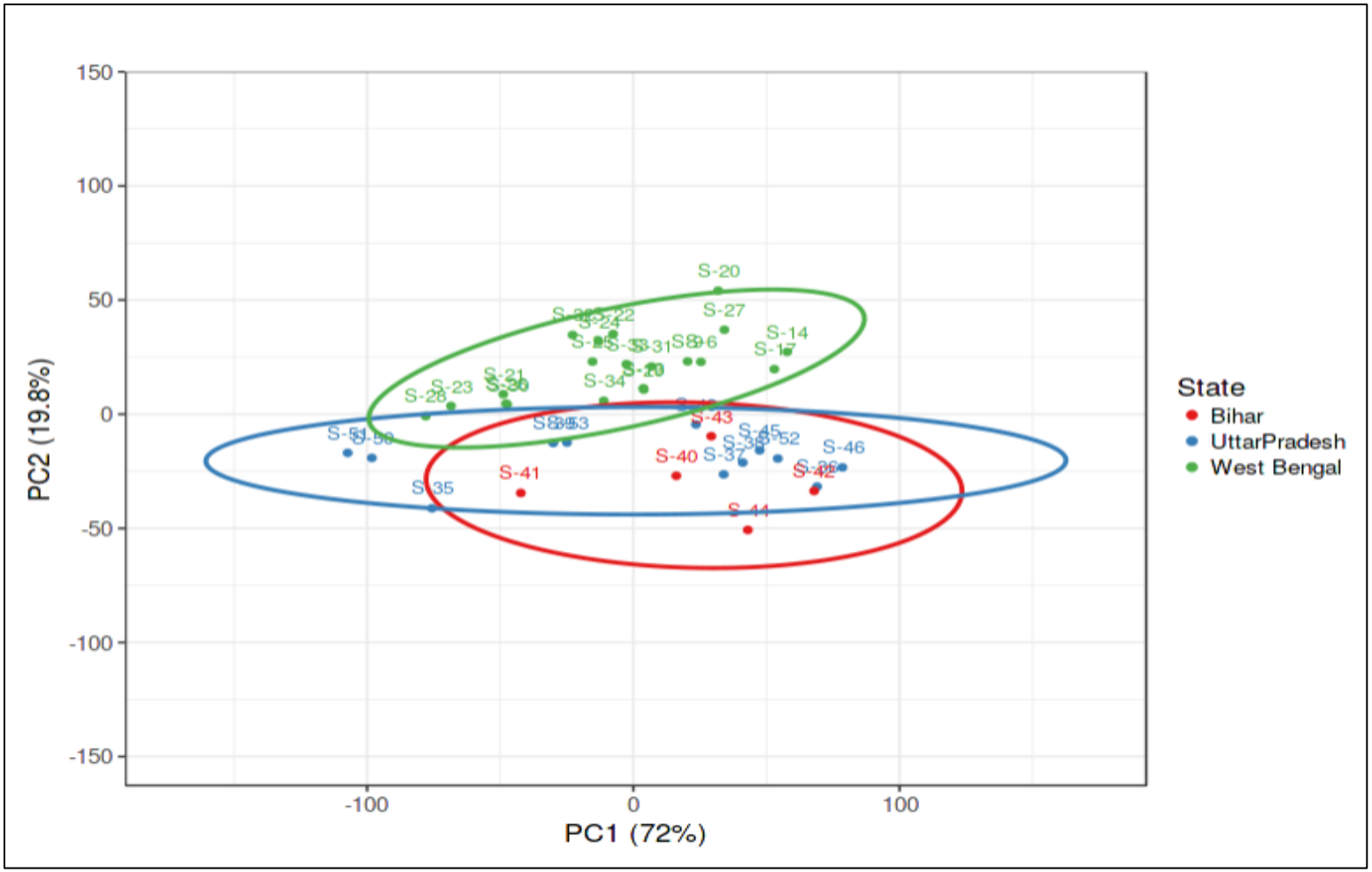


Figure 5. 12 Geographical (among three states) classification of mango on the basis of FTIR spectral data

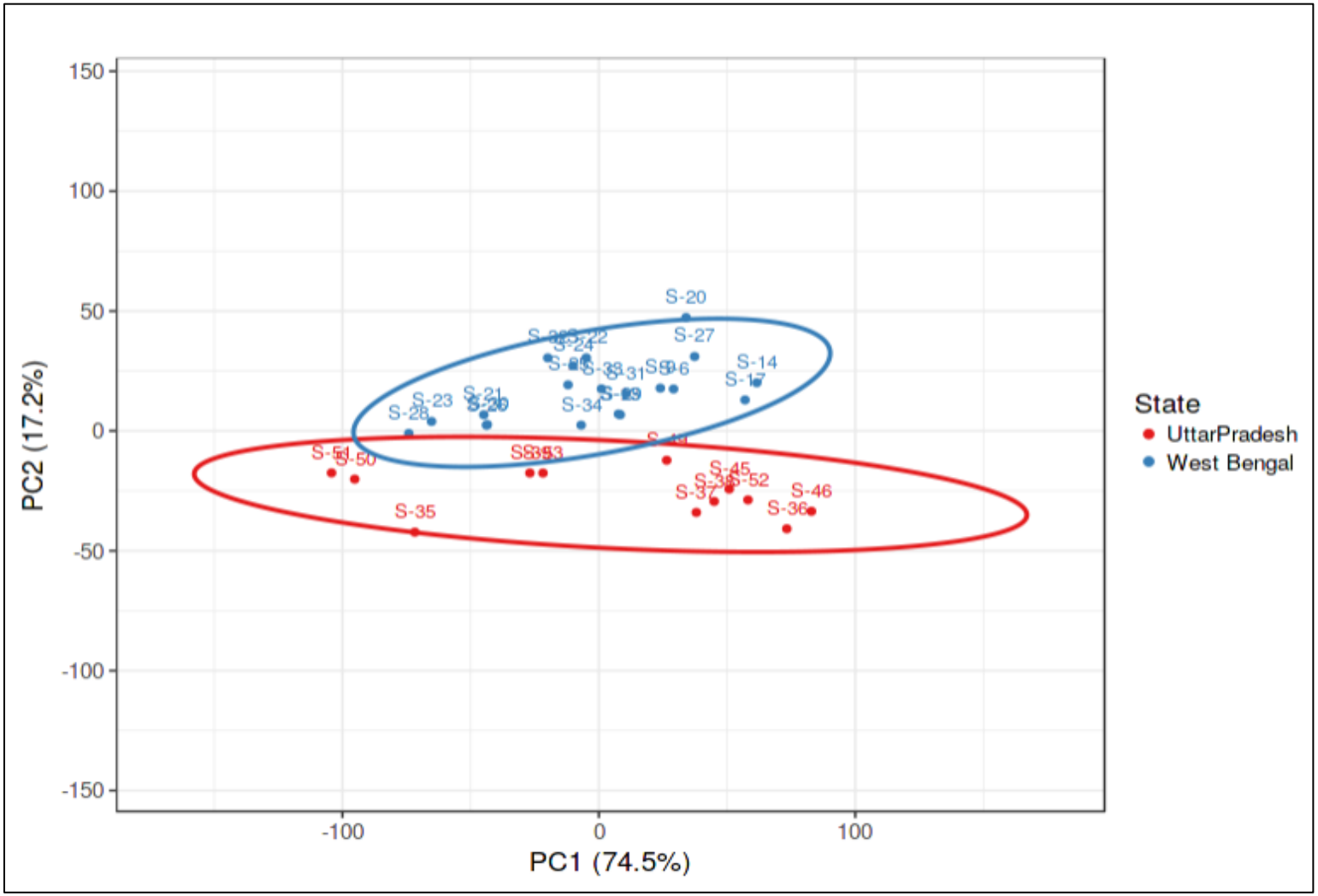


Figure 5. 13 Geographical (among two states) classification of mango on the basis of FTIR spectral data

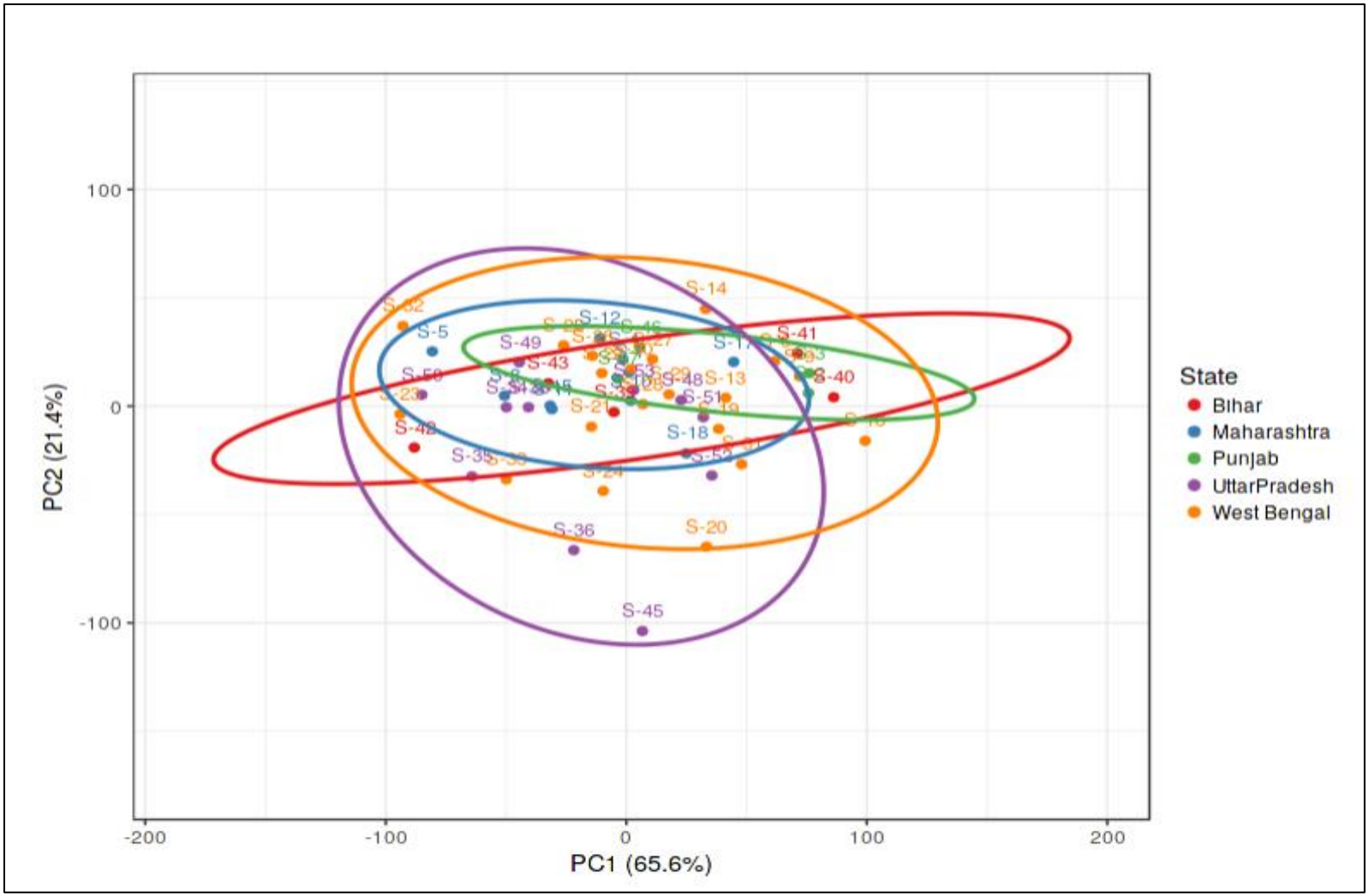


Figure 5. 14 Geographical (among five states) classification of mango on the basis of NIR spectral data

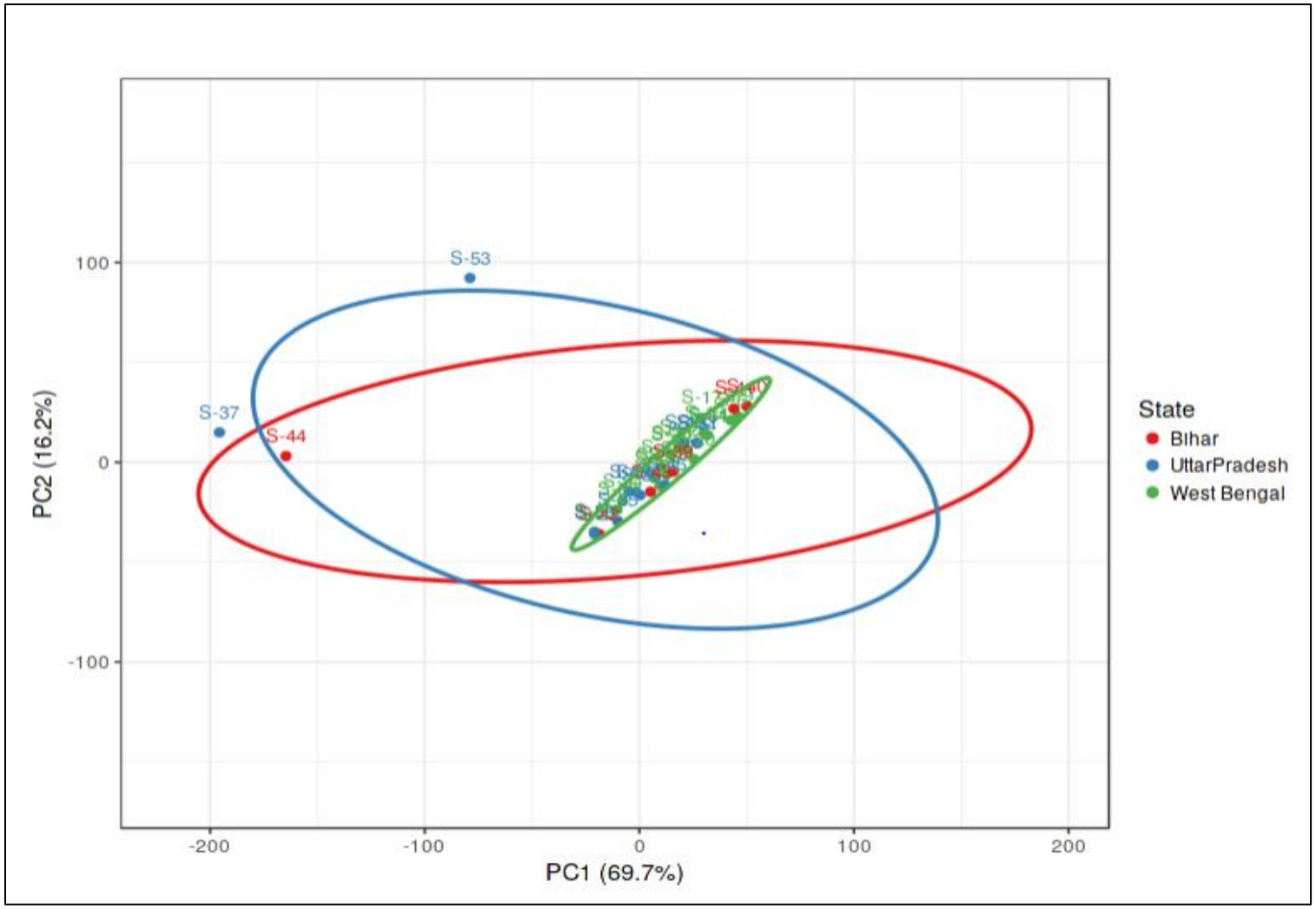


Figure 5. 15 Geographical (among three states) classification of mango on the basis of NIR spectral data

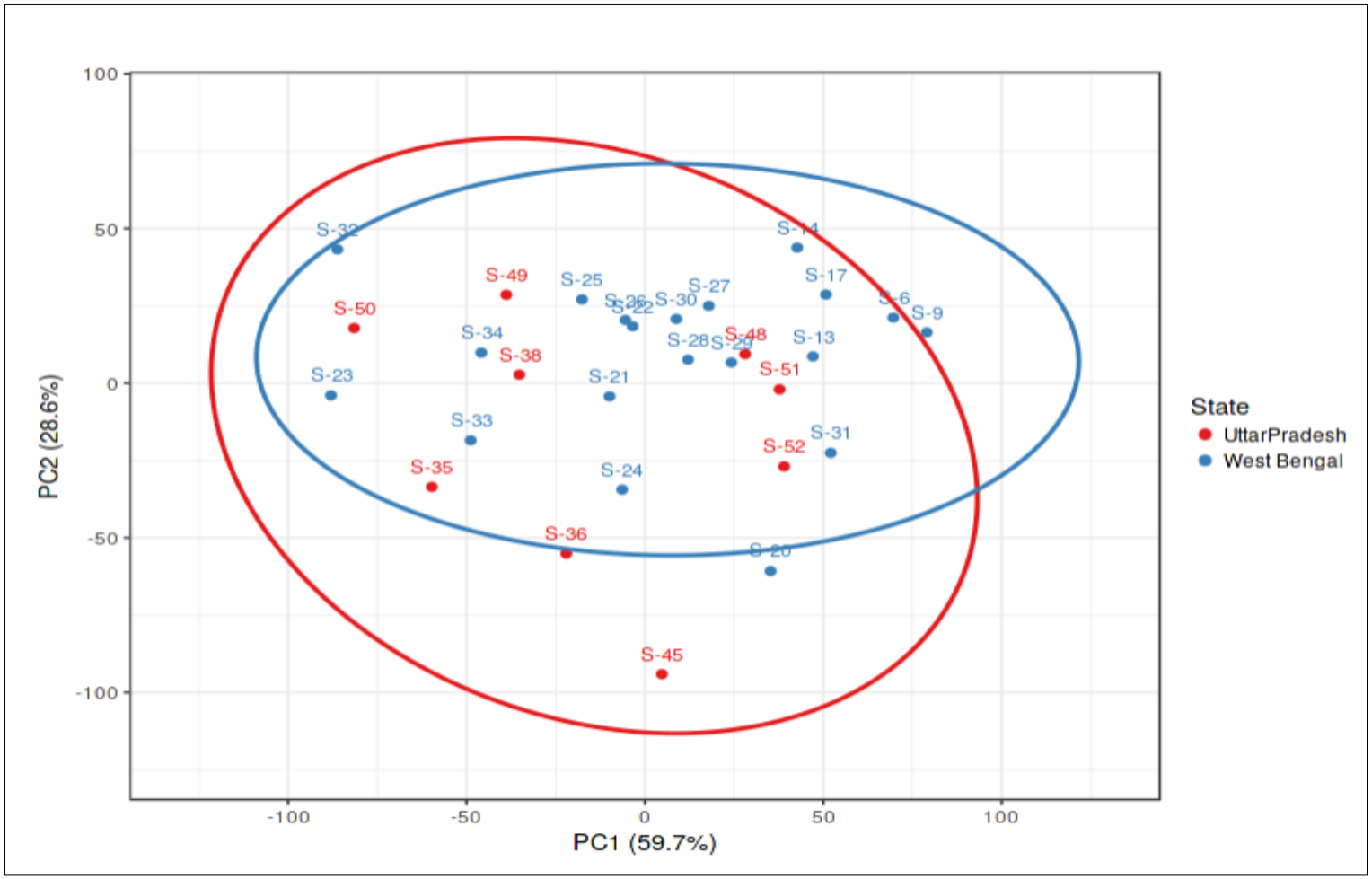


Figure 5. 16 Geographical (among two states) classification of mango on the basis of NIR spectral data

5.4. Conclusion

Quality parameters can be checked with the presence of organic molecules in any sample. Present research was done to check the presence of those organic molecules in mango pulp. Infrared (IR) active molecules have been analyzed with the radiations. IR spectra were recorded in two parts that are near infrared and Fourier Transform infrared spectra. Near infrared and Fourier Transform infrared spectroscopy has been applied to check the presence of organic molecules in the mango. This is also a primary data to estimate the quality parameters of mango pulp along with the chemometrics. PCA was employed using Singular value decomposition algorithm to classify the mango samples on basis of origin. Samples were taken from five different states of India these are Bihar, Maharashtra, Punjab, Uttar Pradesh and West Bengal. The results have concluded that the classification of mango on basis of origin is hardly possible for the total datasets. But after reduction of some data sets, the samples of Uttar Pradesh and West Bengal were classified in a better way. But, among all the parameters that physical attribute, chemical attribute and spectral (NIR and FTIR) attributes, the samples were classified with the FTIR data of the samples of west Bengal and Uttar Pradesh. So, these findings can conclude that the varieties of mangoes are hardly different in nature with each other.

Reference

1. D. Yadav and S. P. Singh, *J. Pharmacogn. Phytochem.*, **6**, 1257-1262 (2017).
2. M. Z. Shafique, M. Ibrahim, M. O. H. Helali and S. K. Biswas, *Bangladesh J. Sci. Ind. Res.*, **41**, 101-108 (2006).
3. U. P. Singh, D. P. Singh, M. Singh, S. Maurya, J. S. Srivastava, R. B. Singh, S. P. Singh, *Int. J. Food Sci. Nutr*., **55**, 163-169 (2004).
4. C. Lizada, "Mango" in *Biochemistry of Fruit Ripening*, edited by G. B. Seymour, J. E. Taylor, G. A. Tucker, (Springer, Dordrecht, 1993), PP. 255-271.
5. I. A. Rajwana, I. A. Khan, A. U. Malik, B. A. Saleem, A. S. Khan, K. Ziaf, R. Anwar and M. Amin, *Int. J. Agric. Biol*, **13**, 151-158 (2011).
6. T. Yantarasri, S. Ben-Yehoshua, V. Rodov, W. Kumpuan, J. Uthaibutra and J. Sornsrivichai, *Postharvest Physiol. Fruit*. **398**, 81-92 (1994).
7. S. N. Jha, A. R. Kingsly and S. Chopra *J. Food Eng*., **72**, 73-76, (2006).
8. T. Osadare, O. Koyenikan, F. Akinola, *Asian Food Sci. J.*, **10**, 1-8 (2019).
9. J. C. Ehiem, K. J. Simonyan, *Int. Agrophys*., **26**, 95-98 (2012).
10. G. Reich, *Adv. Drug Deliv. Rev*., **57**, 1109-1143 (2005).
11. M. Manley, *Chem. Soc. Rev.*, **43**, 8200-8214 (2014).
12. C. Kumaravelu and A. Gopal, "A review on the applications of Near-Infrared spectrometer and Chemometrics for the agro-food processing industries", in IEEE Technological Innovation in ICT for Agriculture and Rural Development (TIAR)-2015, 8-12.
13. D. A. Burns and E. W. Ciurczak, *Handbook of near-infrared analysis*. (CRC press, 2007).
14. I. Jolliffe, I., *J. Multivar. Anal*., **188**, 104820 (2022).
15. H. Abdi, and L. J. Williams, Wiley *Interdis. Rev. Comput. Stat*., **2**, 433-459 (2010).
16. D. Granato, J. S. Santos, G. B. Escher, B. L. Ferreira and R. M. Maggio, *Trends Food Sci. Technol.*, **72**, 83-90 (2018).
17. A. B. Donald and W. C. Emil, *Handbook of Near-Infrared Analysis (Practical Spectroscopy)*, (CRC Press, 2007).

Chapter-6: Regression Model Development

Chapter-6

Regression Model Development

This chapter is dedicated on the discussion about the statistical model development by performing the regression analysis. Regression model is important to estimate all the quality parameters of mango using a green and nondestructive way.

6.1. Introduction

Regression model is the mathematical (statistical) approach to enhance the quality prediction of any organic products [1]. Food and beverage industries is now a days having a leading role for enhancing the economy of the system. Thus, the quality control and food safety measurements become very popular in recent days [2]. Development of a regression model is not possible without the NIR spectroscopy. The spectroscopic data along with the chemmometric study develop the regression model for quality estimation of any food product [3]. All the external and internal properties of any food products can be characterized with the analytical methods. AOAC and AACC are the majorly followed analytical methods to estimate the internal and external quality of any food products. But, these methods are destructive in nature and involved many hazardous chemicals. The long process of analysis time and high cost investment are also one major disadvantage of these methods. To reduce the cost, time and labour involvement Norris and Hart in 1900 have introduced a non-destructive method for quality analysis of organic products [4]. Near infrared spectroscopy has been proved itself a promising technology for non-destructive, robust, non-invasive and non-hazardous method from last two decades [5].

There are some disadvantages too along with the usefulness of near infrared spectroscopy. This technology cannot predict the values of determination for calibration and validation. It needs the prior information about the analytical results to develop the model. Also to predict a good model with better accuracy a large number of dataset is needed for the analysis. As, the near infrared spectroscopy has shown the bands of

combination and overtones of IR active molecules (C-H, S-H, N-H, O-H etc.), the emerging signals have reduced its intensity. To maintain the intensity of the spectra the bands has overlapped with each other. This overlapping makes the signal non-productive and non-useful for the prediction of statistical model. This kind of undesired errors can be occur due to measurements, different external attributes and state of the sample. To make the spectra more productive and accurate many preprocessing treatments have used. Different preprocessing strategies can be applied to remove these kinds of errors. Among them MSC, polynomial baseline correction, First and second order derivative (Savitzky-Golay), standard normal variate (SNV) are important and used in most of the cases for NIR data. Smoothing eliminates the unwanted noise of the spectral signal, Savitzky-Golay derivative (first and second order both) modifies the spectra when the crust and trough, a scattering occurs due the several particle size (physical error) and the scattering can be removed with MSC, SNV also reduce the scattering effect of the spectra and eliminate the variation of slope. Basically by using all the preprocessing strategies the raw spectra can be modified and finally the predictability of the statistical regression model results more accurate [6].

The quantitative estimation of the food products can be verified by performing some analytical approaches namely PLS, PCR and MLR. Among these PLS and PCR are more useful and have the research trend. The acceptance of MLR is less because of its less predictive nature. The errors during the calibration can occur more in this case. MLR can be efficiently applied for small scale of datasets. PCR is used perform PCA first. Among all these methods PLS is one of most popular and commonly used method for the calibration modeling. In a comparison of PLS and PCR method PLS is more useful because of the time effectiveness. PLS algorithm calculates the factors generally in one time during the regression, whereas PCR firstly perform PCA with singular value decomposition (SVD) on all the principal components of total dataset before the regression. So, large dataset can give the result with less time by applying PLS. PLS also deliver the more accurate predictive model as compare to PCR. The initial factors in case of PLS are mostly correlated with the reference values so, it produces more accurate and

precise result. At the same time PCR calculates the principal components (factors) with a descending order of the total variance [7].

6.2. Materials and Method

6.2.1. Sample Collection

Fully ripe mango of different Indigenous varieties was collected from the different mango farms of five states of India, namely, Bihar, Maharashtra, Uttar Pradesh, Punjab and West Bengal. Collected fruits were brought to the Advanced Analytical Testing Laboratory, Kolkata, West Bengal for chemical analysis. The fruits were then washed properly by distilled water to remove the heat, external dirt, dust, pollutants and the pesticide residues.

6.2.2. Sample Preparation

Fifty two different mango samples were sorted and marked. Each mango sample was peeled and the pulp was extracted. The kernel was separated and Lifelong Power pro LLMG02 mixer grinder was used to prepare homogeneous pulp. Pulp samples were then packed into different labeled zip lock pouch bags and finally stored at 4 $^{\circ}$C in refrigerator until future analysis.

6.2.3. Proximate Analysis

The proximate analysis of the homogenized mango pulp was taken with the AOAC standard method. Moisture, crude fiber, crude protein and crude fat were determined with Hot air oven method, Acid-alkali digestion method, Kjeldahl method and Soxhlet method, respectively. pH content and TSS values was determined with the pH meter and hand refractometer. Results were recorded in the triplicate and the mean value was taken. Carbohydrate content was measured with the difference method (Chapter-3).

6.2.4. Regression Model Development

The Near infrared spectral data was collected with the NIR FOSS2500 spectrometer between the spectral ranges of 700-2500 nm. The regression model was developed by importing the NIR spectral data to the data analysis software Unscreambler (CAMO AS, Trondheim, Norway, version 10.5.1). Multivariate analysis was performed on the spectral and chemical data of the mango samples to estimate the nutritional parameters. Raw data

was plotted and visually non-similar data was identified as the outliers and removed from the data set. Carbohydrate, crude fat, crude protein, crude fat, moisture, pH and TSS parameter of total 51 samples were taken for the calibration and validation analysis.

PLS regression and PCR method were used to calibrate and validate the data set. Some pre-processing methods, MSC and second order derivative (Using Savitzky Golay method with second order polynomial and 3 smoothing points) were applied on the original raw spectral data to improve the model predictability. In order to search the minimum range of wavelength for good prediction model for sweetness of mango the total range of wavelength was divided into 18 groups. Each group contained a range of 100 nm wavelengths and PLS, iPLS and siPLS were applied on each group. The best prediction model was introduced with the Coefficient of determination for calibration, Coefficient of determination for validation, Root mean square error for calibration and Root mean square error for validation values. The plot between reference and predicted values of pH and TSS was presented to know the actual predictability using NIR spectroscopy in a non-destructive way.

6.3. Results and Discussion

To predict the chemical parameters of the mango, prediction of the all parameters (carbohydrate, crude fat, crude protein, crude fiber, moisture, pH and TSS) were done with non-destructive method. For nondestructive evaluation two particular approaches has been used The values of the Coefficient of determination for calibration (R_c^2), Coefficient of determination for validation (R_v^2), RMSEC and RMSEV judged the quality of the prediction model. A good quality regression model has the lower values of RMSEC and RMSEV and higher values of R_c^2 and R_v^2. PLS regression was applied to the raw data for the prediction. Particularly two different pre-processing methods were suitable and applied to the raw spectral data to get the more accurate result. MSC modified the original spectra by bringing all different spectra of different samples more nearly. Derivative computed the overlapping crust and trough of each spectrum and in the present case second order derivative proved as efficient to modify the spectra. After applied MSC and second derivative the result in terms of coefficients and errors were

modified for prediction of every qualitative parameter.

6.3.1. Partial Least Square Regression and Principal Component Regression

The estimation of the quality parameters (moisture, crude fat, crude protein, crude fiber, carbohydrate, pH and TSS) of mango pulp was done with the PLS regression method using spectral and reference data. The original spectra of the raw NIR data of total 51 different mango pulp samples have been presented in Figure 5.1 from 700 nm to 2500 nm with the near infrared spectrometer and have shown some bands and peaks at different wavelengths which have indicated the presence of different organic molecules in the mango pulp sample. Figure 5.2 represents the spectra of one sample highlighting the bands / peaks in the spectra. The values of R_c^2, R_v^2, RMSEC and RMSEV have judged the quality of the prediction model. A good quality regression model has the lower values of RMSEC and RMSEV and higher values of R_c^2 and R_v^2. PLS regression was applied to the raw data for the prediction. The values of R_c^2, R_v^2, RMSEC and RMSEV for moisture content, crude fat, crude protein, crude fiber, carbohydrate, pH and total soluble solid have been presented in Table 6.1, 6.9, 6.17, 6.25, 6.33, 6.41 and 6.49. These tables also represent the effect of some pre-processing strategies on the values of coefficients and errors. It has been clearly predicted from these tables (Table 6.1, 6.9, 6.17, 6.25, 6.33, 6.41 and 6.49) that the effect of MSC and second order derivative is best for this in terms of higher values of coefficient of determinations and lower values of root mean square errors. Particularly two different pre-processing methods were suitable and have been applied to the raw spectral data to get the more accurate result. Mango pulp is not homogenous and due to this when NIR radiation has illuminated on the sample, it results into a scattering error in final spectra. MSC has eliminated these scattering errors from the spectra and modified the results in more accurate way. On the other hand derivative has been computed the overlapping crust and trough of each spectrum and in the present case second order derivative has been proved efficient to modify and reduces the errors occurred due to the overlapping of crust and trough of the spectra. Maximum of 7 factors of PLS has been taken in the present study to check the effective value and factor 7 has been shown the best results for every case. For principal component regression method

the same pre-processing method was used to determine the results. PCR also has been taken maximum 7 factors to check the effective value and in this case different factors showed importance for every parameter (Table 6.8, 6.16, 6.24, 6.32, 6.40, 6.48 and 6.56)

6.3.2. Interval Partial Least Square Regression and Synergic Partial Least Square Regression

Wavelength selection method was used to determine the values of coefficients and errors more accurately. The final objective of current research is to obtain a robust regression model to predict the moisture content of mango pulp with minimum wavelengths. The wavelength reduction methods used for analysis were - iPLS and siPLS. A bar diagram has elaborated the values of coefficient of determination and root mean square error for both calibration and validation for total range and interval wavelength ranges (Figure 6.2, 6.5, 6.8, 6.11, 6.14, 6.17 and 6.20).

For iPLS, the total range of wavelength has been divided into some equal intervals of 600 nm, 300 nm and 200 nm of equal width. The wavelength groups with 600 nm intervals were taken as 700-1300 nm, 1300-1900 nm and 1900-2500 nm. For all the cases, R_c^2, R_v^2, RMSEC and RMSEV values were determined with PLS regression method. The values were determined for all the factors from 2-7, to check the variability of the values of coefficient and errors.

Further wavelength selection process was done by siPLS method. This method selected those ranges that have given the best results in iPLS and were compiled together, irrespective of their width. Along with the PLS regression PCR method was also applied to predict the same parameters (R_c^2, R_v^2, RMSEC and RMSEV). For PCR, the regression method was applied for the total range (700-2500 nm). The results were obtained for the raw data and pre-processed data as well. The same pre-processing approaches (MSC, 2^{nd} order derivative and MSC+2^{nd} order derivative) were applied to the NIR spectral data. For PCR analysis it was observed in Table Table 6.8, 6.16, 6.24, 6.32, 6.40, 6.48 and 6.56 that the raw data have the best values of R_c^2, R_v^2, RMSEC and RMSEV. The preprocessed data has given low values of coefficients and high values of root mean square errors as compared to PLS regression model.

6.3.3. Estimation of quality parameter

- **Moisture:**

In original (without preprocessing) spectra, the value of R_c^2 is 0.6028, which has increased to 0.6460, 0.9865 and 0.9937 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Similarly, the value of R_v^2 is 0.3975, which has increased to 0.4947, 0.9818 and 0.9907 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Finally the values of coefficients and errors after performing the pre-processing MSC and second order derivative together at the factor 7 have been considered for total wavelength range (700-2500 nm) for interval and synergy interval PLS.

In Table 6.2, 6.3 and 6.4 it has been observed that factor 7 is showing the best result in terms of values. So, it was determined that for all cases factor 7 is significantly important and the further results for other intervals has been evaluated with 7 factors and results are presented in Table 6.5 and 6.6. The results in Table 6.2, 6.3 and 6.4 showed that 700-1300 nm range is not important for the moisture content estimation as it has high error. The important wavelengths for estimation of moisture content have been in the range from 1300-2500 nm.

For PCR, in original spectra the value of R_c^2 is 0.4524, which has changed to 0.3487, 0.3169 and 0.3841 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Similarly, the value of R_v^2 is 0.1180, which has increased to 0.0767, 0.1225 after applying MSC and MSC+ Second order derivative respectively. Due to insufficient data for second order derivative the regression model could not be developed.

Table 6.7 has explained about some selected ranges which has given the finest results in terms of root mean square errors and coefficient of determination values. Among these two segments of Table 6.7 the range 2100-2500 nm has the minimum wavelengths and best results as 0.9569 (R_c^2) and 0.9529 (R_v^2). So the estimation of moisture content can be done with the use of 2100-2500 nm wavelength range with results equivalent to the complete spectrum with a non-destructive approach.

- **Crude Fat:**

In original (without preprocessing) spectra, the value of R_c^2 is 0.4293, which has increased to 0.5242, 0.9604 and 0.9492 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Similarly, the value of R_v^2 is 0.2420, which has increased to 0.3432, 0.9478 and 0.9492 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Finally the values of coefficients and errors after performing the pre-processing MSC and second order derivative together at the factor 7 have been considered for total wavelength range (700-2500 nm) for interval and synergy interval PLS.

In Table 6.10, 611 and 6.12 it has been observed that factor 7 is showing the best result in terms of values. So, it was determined that for all cases factor 7 is significantly important and the further results for other intervals has been evaluated with 7 factors and results are presented in Table 6.13 and 6.14. The results in Table 6.10, 6.11 and 6.12 showed that 700-1300 nm range is not important for the moisture content estimation as it has high error. The important wavelengths for estimation of moisture content have been in the range from 1300-2500 nm.

For PCR, in original spectra the value of R_c^2 is 0.1890, which has changed to 0.2829, 0.3228 and 0.2267 after applying MSC, second order derivative and MSC+ Second order derivative respectively. But, the value of R_v^2 couldn't determine due to insufficient data for MSC and MSC+ Second order derivative and second order derivative.

Table 6.15 has explained about some selected ranges which has given the finest results in terms of root mean square errors and coefficient of determination values. Among these two segments of Table 6.15 the range 1100-1300, 1500-2300 nm has the minimum wavelengths and best results as 0.9961 (R_c^2) and 0.9044 (R_v^2). So the estimation of crude fat content can be done with the use of 1100-1300, 1500-2300 nm wavelength range with results equivalent to the complete spectrum with a non-destructive approach.

- **Crude Protein:**

In original (without preprocessing) spectra, the value of R_c^2 is 0.2937, which has increased to 0.4754, 0.9816 and 0.9822 after applying MSC, second order derivative and

MSC+ Second order derivative respectively. Similarly, the value of R_v^2 couldn't determine for original spectra, but the value has increased from 0.2397 to 0.9740 and 0.9744 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Finally the values of coefficients and errors after performing the pre-processing MSC and second order derivative together at the factor 7 have been considered for total wavelength range (700-2500 nm) for interval and synergy interval PLS.

In Table 6.16, 617 and 6.18 it has been observed that factor 7 is showing the best result in terms of values. So, it was determined that for all cases factor 7 is significantly important and the further results for other intervals has been evaluated with 7 factors and results are presented in Table 6.19 and 6.20. The results in Table 6.16, 6.17 and 6.18 showed that 700-1300 nm range is not important for the moisture content estimation as it has high error. The important wavelengths for estimation of moisture content have been in the range from 1300-2500 nm.

For PCR, in original spectra the value of R_c^2 is 0.1890, which has changed to 0.2829, 0.3228 and 0.2267 after applying MSC, second order derivative and MSC+ Second order derivative respectively. But, the value of R_v^2 couldn't determine due to insufficient data for MSC and MSC+ Second order derivative and second order derivative.

Table 6.15 has explained about some selected ranges which has given the finest results. in terms of root mean square errors and coefficient of determination values. Among these two segments of Table 6.15 the range 1100-1300, 1500-2300 nm has the minimum wavelengths and best results as 0.9961 (R_c^2) and 0.9044 (R_v^2). So the estimation of moisture content can be done with the use of 1100-1300, 1500-2300 nm wavelength range with results equivalent to the complete spectrum with a non-destructive approach.

- **Crude Fiber:**

In original (without preprocessing) spectra, the value of R_c^2 is 0.4997, which has increased to 0.5679, 0.9709 and 0.9712 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Similarly, the value of R_v^2 is 0.2904, which has increased to 0.3648, 0.9582 and 0.9590 after applying MSC, second order derivative

and MSC+ Second order derivative respectively. Finally the values of coefficients and errors after performing the pre-processing MSC and second order derivative together at the factor 7 have been considered for total wavelength range (700-2500 nm) for interval and synergy interval PLS.

In Table 6.24, 6.25 and 6.26 it has been observed that factor 7 is showing the best result in terms of values. So, it was determined that for all cases factor 7 is significantly important and the further results for other intervals has been evaluated with 7 factors and results are presented in Table 6.27 and 6.28. The results in Table 6.24, 6.25 and 6.26 showed that 700-1300 nm range is not important for the moisture content estimation as it has high error. The important wavelengths for estimation of moisture content have been in the range from 1300-2500 nm.

For PCR, in original spectra the value of R_c^2 is 0.2455, which has changed to 0.2472, 0.2221 and 0.2679 after applying MSC, second order derivative and MSC+ Second order derivative respectively. But, the value of R_v^2 is 0.0063 for original spectra but the same value couldn't determine due to insufficient data for MSC and MSC+ Second order derivative and second order derivative.

Table 6.29 has explained about some selected ranges which has given the finest results in terms of root mean square errors and coefficient of determination values. Among these two segments of Table 6.29 the range 1100-1300, 1500-2300 nm has the minimum wavelengths and best results as 0.9961 (R_c^2) and 0.9044 (R_v^2). So the estimation of moisture content can be done with the use of 1100-1300, 1500-2300 nm wavelength range with results equivalent to the complete spectrum with a non-destructive approach.

- **pH Content:**

In original (without preprocessing) spectra, the value of R_c^2 is 0.4779, which has increased to 0.6039, 0.9905 and 0.9913 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Similarly, the value of R_v^2 is 0.2955, but the value has increased from 0.3935, 0.9856 and 0.9868 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Finally the values of coefficients and errors after performing the pre-processing MSC and second order

derivative together at the factor 7 have been considered for total wavelength range (700-2500 nm) for interval and synergy interval PLS.

In Table 6.32, 6.33 and 6.34 it has been observed that factor 7 is showing the best result in terms of values. So, it was determined that for all cases factor 7 is significantly important and the further results for other intervals has been evaluated with 7 factors and results are presented in Table 6.35 and 6.36. The results in Table 6.32, 6.33 and 6.34 showed that 700-1300 nm range is not important for the moisture content estimation as it has high error. The important wavelengths for estimation of moisture content have been in the range from 1300-2500 nm.

For PCR, in original spectra the value of R_c^2 is 0.1697, which has changed to 0.1302, 0.1108 and 0.3356 after applying MSC, second order derivative and MSC+ Second order derivative respectively. But, the value of R_v^2 couldn't determine due to insufficient data for original spectra, MSC and MSC+ second order derivative, but for second order derivative spectra the value of R_v^2 is 0.0390.

Table 6.37 has explained about some selected ranges which has given the finest results. in terms of root mean square errors and coefficient of determination values. Among these two segments of Table 6.37 the range 1100-1300, 1500-2300 nm has the minimum wavelengths and best results as 0.9961 (R_c^2) and 0.9044 (R_v^2). So the estimation of moisture content can be done with the use of 1100-1300, 1500-2300 nm wavelength range with results equivalent to the complete spectrum with a non-destructive approach.

- **Carbohydrate:**

In original (without preprocessing) spectra, the value of R_c^2 is 0.5945, which has increased to 0.6459, 0.9910 and 0.9923 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Similarly, the value of R_v^2 is 0.3866, which has increased to 0.4922, 0.9975 and 0.9892 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Finally the values of coefficients and errors after performing the pre-processing MSC and second order derivative together at the factor 7 have been considered for total wavelength range (700-2500 nm) for interval and synergy interval PLS.

In Table 6.40, 6.41 and 6.42 it has been observed that factor 7 is showing the best result in terms of values. So, it was determined that for all cases factor 7 is significantly important and the further results for other intervals has been evaluated with 7 factors and results are presented in Table 6.43 and 6.44. The results in Table 6.40, 6.41 and 6.42 showed that 700-1300 nm range is not important for the moisture content estimation as it has high error. The important wavelengths for estimation of moisture content have been in the range from 1300-2500 nm.

For PCR, in original spectra the value of R_c^2 is 0.4402, which has changed to 0.2865, 0.2836 and 0.3274 after applying MSC, second order derivative and MSC+ Second order derivative respectively and the value of R_v^2 is 0.0674, 0.0323 and 0.0283 for original spectra, MSC and MSC+ Second order derivative but the same value couldn't determine due to insufficient data for second order derivative.

Table 6.45 has explained about some selected ranges which has given the finest results. in terms of root mean square errors and coefficient of determination values. Among these two segments of Table 6.45 the range 1100-1300, 1500-2300 nm has the minimum wavelengths and best results as 0.9961 (R_c^2) and 0.9044 (R_v^2). So the estimation of moisture content can be done with the use of 1100-1300, 1500-2300 nm wavelength range with results equivalent to the complete spectrum with a non-destructive approach.

- **Total Soluble Solid:**

In original (without preprocessing) spectra, the value of R_c^2 is 0.5945, which has increased to 0.6459, 0.9910 and 0.9923 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Similarly, the value of R_v^2 is 0.3866, which has increased to 0.4922, 0.9875 and 0.9892 after applying MSC, second order derivative and MSC+ Second order derivative respectively. Finally the values of coefficients and errors after performing the pre-processing MSC and second order derivative together at the factor 7 have been considered for total wavelength range (700-2500 nm) for interval and synergy interval PLS.

In Table 6.40, 6.41 and 6.42 it has been observed that factor 7 is showing the best result in terms of values. So, it was determined that for all cases factor 7 is significantly

important and the further results for other intervals has been evaluated with 7 factors and results are presented in Table 6.43 and 6.44. The results in Table 6.40, 6.41 and 6.42 showed that 700-1300 nm range is not important for the moisture content estimation as it has high error. The important wavelengths for estimation of moisture content have been in the range from 1300-2500 nm.

For PCR, in original spectra the value of R_c^2 is 0.1527, which has changed to 0.1942, 0.1397 and 0.1297 after applying MSC, second order derivative and MSC+ Second order derivative respectively. But, the value of R_v^2 couldn't determine due to insufficient data for MSC and MSC+ Second order derivative and second order derivative.

Table 6.53 has explained about some selected ranges which has given the finest results. in terms of root mean square errors and coefficient of determination values. Among these two segments of Table 6.53 the range 1100-1300, 1500-2300 nm has the minimum wavelengths and best results as 0.9961 (R_c^2) and 0.9044 (R_v^2). So the estimation of moisture content can be done with the use of 1100-1300, 1500-2300 nm wavelength range with results equivalent to the complete spectrum with a non-destructive approach.

A comparison two different multivariate approaches has been done. The values of coefficient of determination and root mean square error for calibration and validation were not so much effective for the rapid determination of the moisture content with PCR method. The estimation of moisture content in mango pulp was done with PLS regression method in a better way using iPLS and siPLS approach.

Table 6. 1 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for moisture with Partial Least Square regression

Preprocessing techniques	**Factors**	**R_c^2**	**RMSEC**	**R_v^2**	**RMSEV**
Original spectra	2	0.3589	0.0337	0.2659	0.0361
	3	0.4429	0.0314	0.3222	0.0347
	4	0.4491	0.0313	0.2406	0.0367
	5	0.5014	0.0297	0.2834	0.0357
	6	0.5741	0.0275	0.3756	0.0333
	7	0.6028	0.0265	0.3975	0.0327
MSC	2	0.1083	0.0398	NA	0.0429
	3	0.2948	0.0354	0.1506	0.0388
	4	0.3946	0.0328	0.2504	0.0365
	5	0.4624	0.0309	0.3156	0.0348
	6	0.5430	0.0285	0.3909	0.0329
	7	0.6460	0.0250	0.4947	0.0299
2^{nd} order derivative	2	0.6781	0.0239	0.6449	0.0251
	3	0.8184	0.0179	0.7880	0.0194
	4	0.9040	0.0130	0.8821	0.0143
	5	0.9638	0.0080	0.9559	0.0088
	6	0.9865	0.0048	0.9818	0.0056
	7	0.9942	0.0032	0.9915	0.0038
MSC + 2^{nd} order derivative	2	0.6792	0.0238	0.6454	0.0251
	3	0.8186	0.0179	0.7879	0.0194
	4	0.9083	0.0127	0.8885	0.0140
	5	0.9660	0.0077	0.9582	0.0086
	6	0.9853	0.0051	0.9804	0.0058
	7	0.9937	0.0033	0.9907	0.0040

Table 6. 2 interval Partial Least Square regression model evaluation of moisture within the range of 700-1300 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.2430	0.0367	0.1459	0.0389
3	0.2987	0.0353	0.8866	0.0402
4	0.3342	0.0344	0.1038	0.0399
5	0.3576	0.0338	0.1293	0.0393
6	0.4290	0.0318	0.1098	0.0397
7	0.4668	0.0307	0.1752	0.0383

Table 6. 3 interval Partial Least Square regression model evaluation of moisture within the range of 1300-1900 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.4380	0.0316	0.3519	0.0339
3	0.5677	0.0277	0.4848	0.0302
4	0.7862	0.0195	0.7231	0.0221
5	0.8968	0.0062	0.8562	0.0159
6	0.9427	0.0100	0.9183	0.0120
7	0.9678	0.0075	0.9519	0.0092

Table 6. 4 interval Partial Least Square Regression model evaluation of moisture within the range of 1900-2500 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.6356	0.0254	0.5722	0.0275
3	0.7589	0.0207	0.7245	0.0221
4	0.8592	0.0158	0.8350	0.0171
5	0.9155	0.0122	0.8981	0.0134
6	0.9496	0.0094	0.9348	0.0107

7	0.9657	0.0078	0.9538	0.0090

Table 6. 5 interval Partial Least Square Regression model evaluation of moisture with the interval of 300 nm

Wavelength	**R_c^2**	**RMSEC**	**R_v^2**	**RMSEV**
700-1000	0.3409	0.0342	0.2699	0.0360
1000-1300	0.3878	0.0330	0.2907	0.0355
1300-1600	0.4288	0.0318	0.3085	0.0350
1600-1900	0.4732	0.0306	0.3414	0.0342
1900-2200	0.5252	0.0290	0.3478	0.0340
2200-2500	0.5736	0.0275	0.3658	0.0335

Table 6. 6 interval Partial Least Square Regression model evaluation of moisture with the interval of 200 nm

Wavelength	**R_c^2**	**RMSEC**	**R_v^2**	**RMSEV**
700-900	0.5509	0.0282	0.3960	0.0327
900-1100	0.4203	0.0321	0.1689	0.0384
1100-1300	0.7207	0.0222	0.6028	0.0265
1300-1500	0.7786	0.0198	0.6726	0.0241
1500-1700	0.8365	0.0170	0.7959	0.0190
1700-1900	0.7811	0.0197	0.6871	0.0235
1900-2100	0.7680	0.0203	0.6994	0.0231
2100-2300	0.8239	0.0176	0.7396	0.0215
2300-2500	0.8705	0.0151	0.8325	0.0172

Table 6. 7 synergic interval Partial Least Square Regression model evaluation of moisture with some selected wavelengths

Wavelength	**R_c^2**	**RMSEC**	**R_v^2**	**RMSEV**
1500-1700, 2100-2500	0.9703	0.0017	0.9589	0.0521

1300-1500, 2300-2500	0.9569	0.0093	0.9529	0.0153

Table 6. 8 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for moisture with Principle Component Regression method

Preprocessing techniques	**Factors**	**R_c^2**	**RMSEC**	**R_v^2**	**RMSEV**
Original spectra	2	0.2681	0.0360	0.1671	0.0384
	3	0.2748	0.0359	0.0585	0.0409
	4	0.4504	0.0317	0.2188	0.0372
	5	0.4504	0.0312	0.1564	0.0387
	6	0.4523	0.0316	0.1564	0.0387
	7	0.4524	0.0319	0.1180	0.0396
MSC	2	0.0175	3.6569	NA	3.9604
	3	0.0569	3.5828	NA	5.4636
	4	0.0986	3.5028	NA	5.4385
	5	0.2002	3.2995	NA	4.7518
	6	0.3277	3.0250	0.0640	3.5694
	7	0.3487	2.9774	0.0767	3.5450
2^{nd} order derivative	2	0.0477	3.6689	NA	3.9237
	3	0.0562	3.6525	NA	3.9740
	4	0.1369	3.4929	NA	3.9294
	5	0.1382	3.4903	NA	4.0955
	6	0.1747	3.4154	NA	4.1091
	7	0.3169	3.1074	NA	5.1022
MSC + 2^{nd} order derivative	2	0.1825	3.3993	0.0898	3.5869
	3	0.1947	3.3773	0.0580	3.6491
	4	0.2303	3.2983	0.0339	3.6954
	5	0.2981	3.1497	0.0860	3.5943
	6	0.3808	2.9585	0.1512	3.4639
	7	0.3841	2.9506	0.1225	3.5219

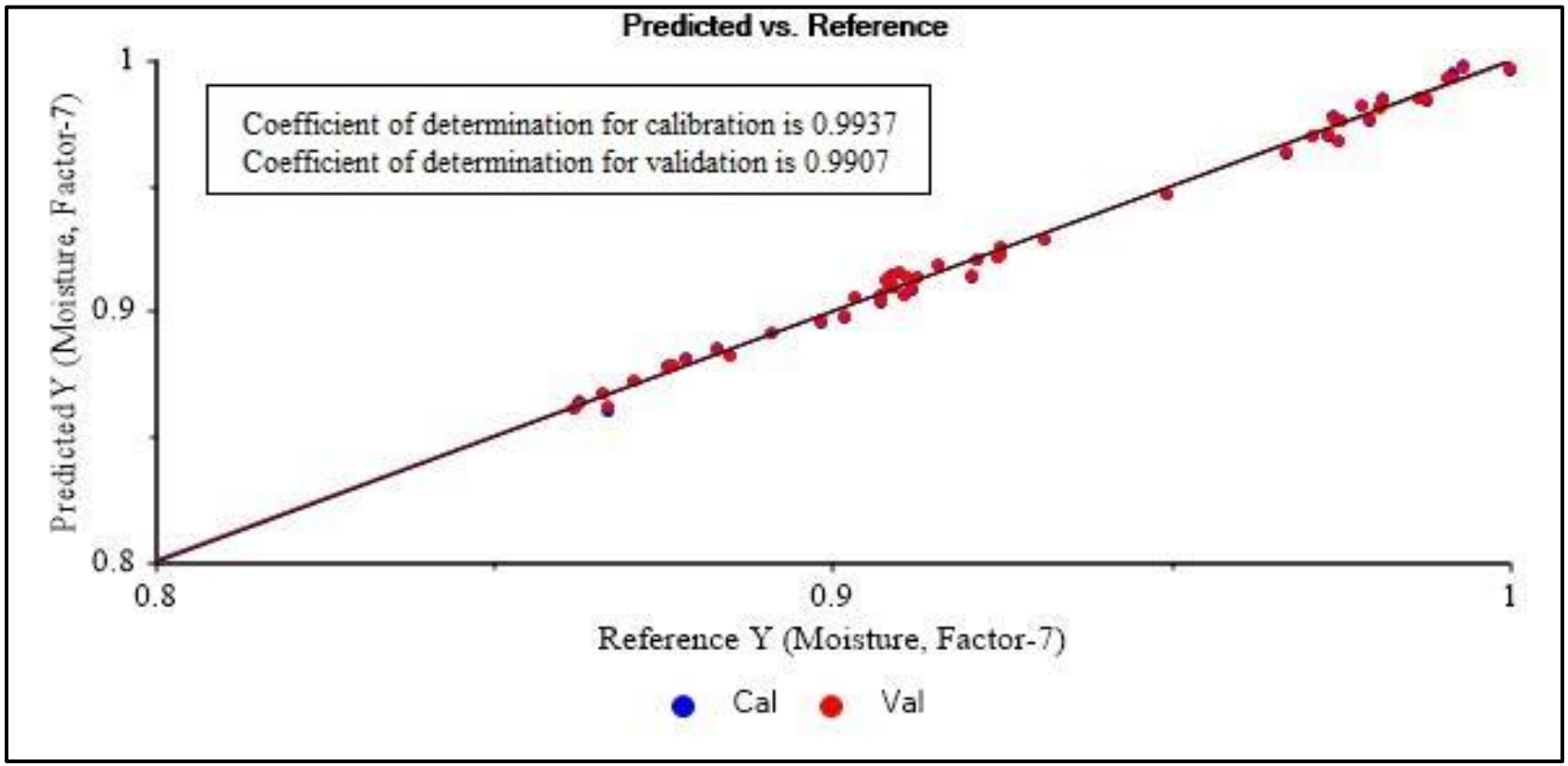


Figure 6. 1 Scatter plot of mango varieties for moisture content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation with Partial Least Square regression

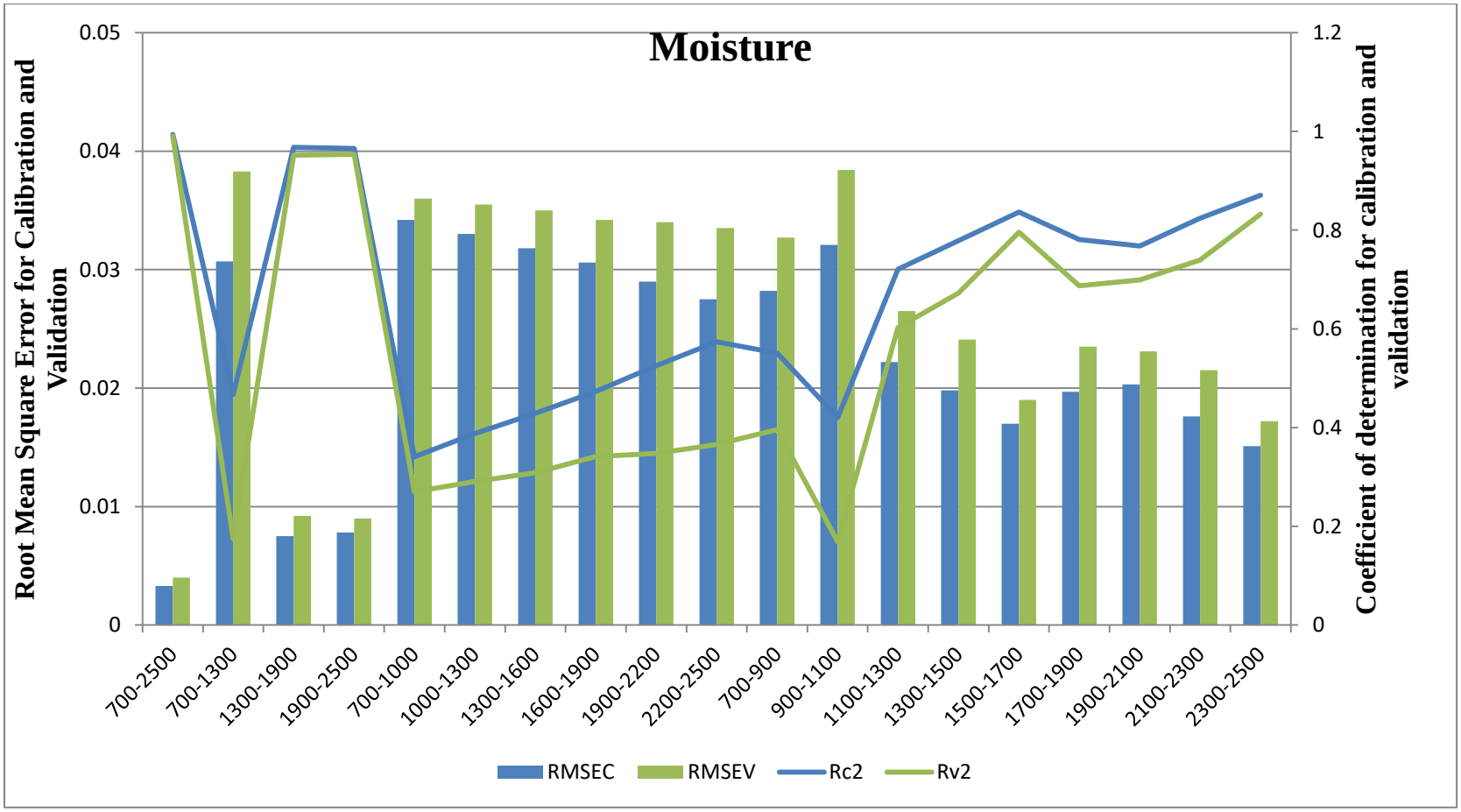


Figure 6. 2 Variation of regression coefficient and root mean square error for calibration and validation with different wavelength range

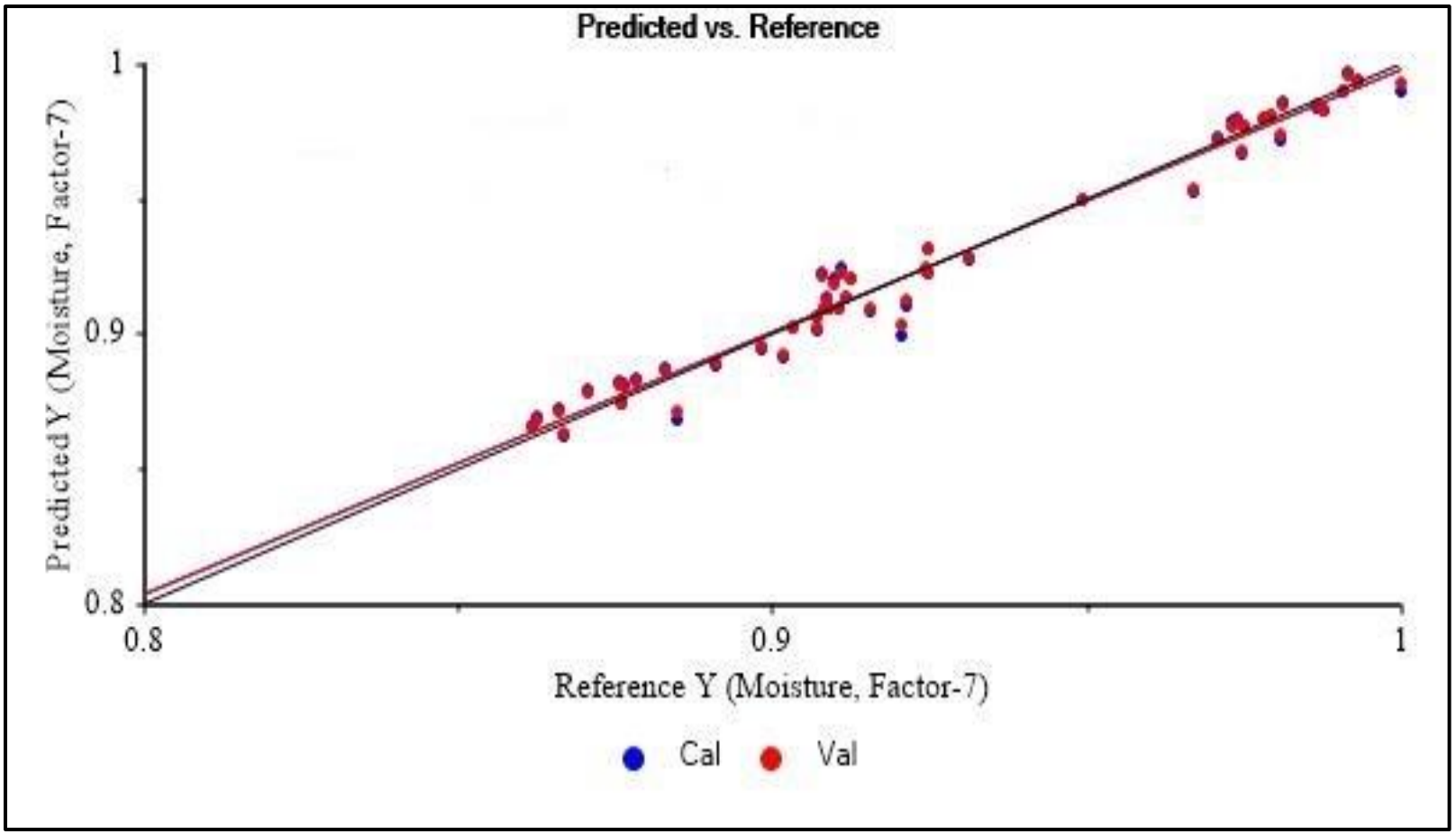


Figure 6. 3 Scatter plot of mango varieties for moisture content synergic interval Partial Least Square regression in selected wavelength range for calibration and validation

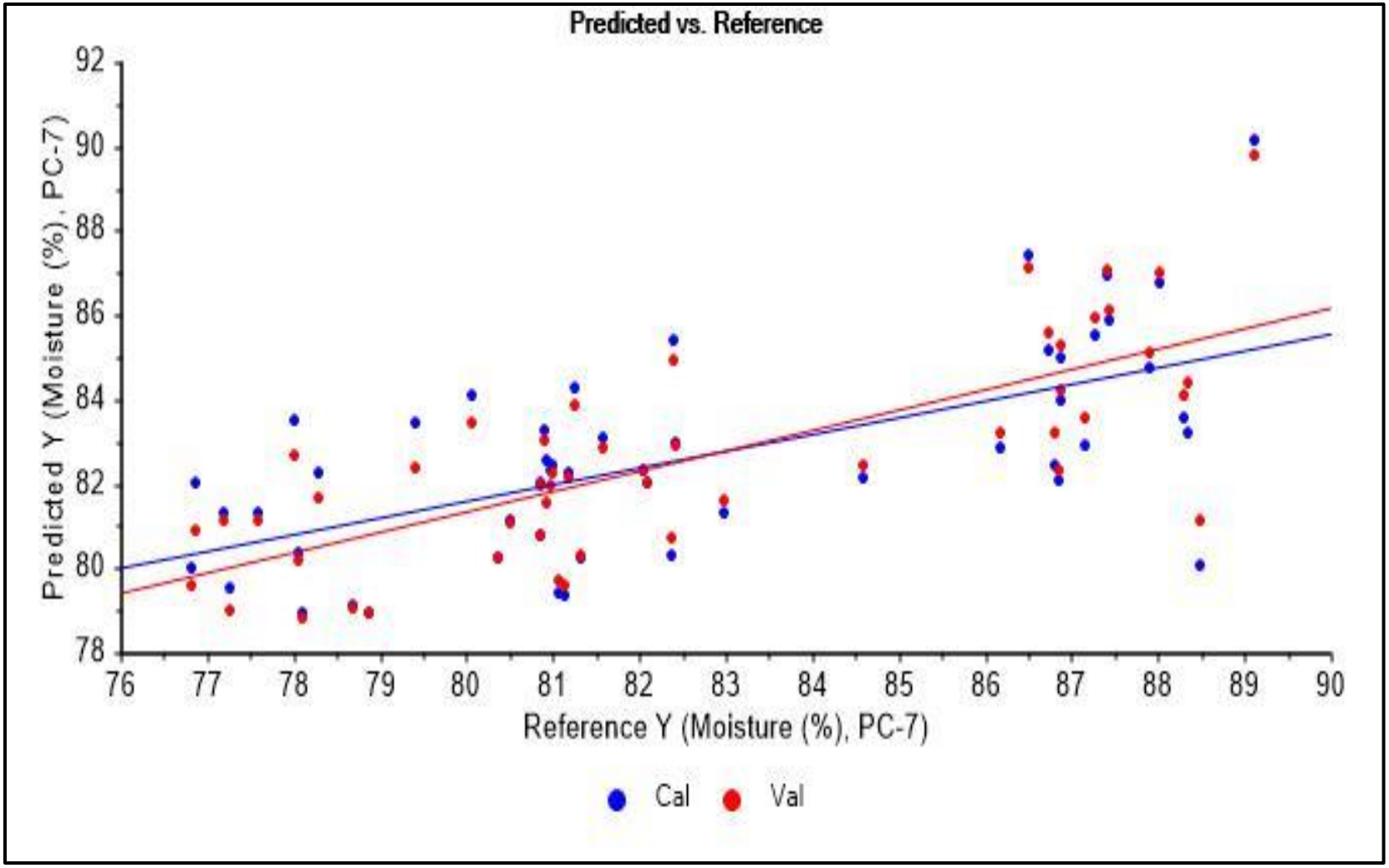


Figure 6. 4 Scatter plot of mango varieties for moisture content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation with Principal Component Regression

Table 6. 9 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for crude fat with Partial Least Square regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.1230	0.2204	0.0120	0.2339
	3	0.1361	0.2187	NA	0.2382
	4	0.1743	0.2138	NA	0.2369
	5	0.2959	0.1974	0.1034	0.2228
	6	0.3480	0.1900	0.1655	0.2149
	7	0.4293	0.1777	0.2420	0.2048
MSC	2	0.1132	0.2216	NA	0.2362
	3	0.2533	0.2033	0.1287	0.2196
	4	0.3183	0.1943	0.1903	0.2117
	5	0.3436	0.1906	0.1961	0.2110
	6	0.3890	0.1839	0.1892	0.2119
	7	0.5242	0.1623	0.3432	0.1907
2^{nd} order derivative	2	0.6249	0.1441	0.5770	0.1530
	3	0.7297	0.1223	0.6815	0.1328
	4	0.8138	0.1015	0.7630	0.1145
	5	0.9035	0.0730	0.8774	0.0823
	6	0.9332	0.0608	0.9135	0.0692
	7	0.9604	0.0468	0.9478	0.0537
MSC + 2^{nd} order derivative	2	0.6233	0.1444	0.5742	0.1535
	3	0.7260	0.1231	0.6762	0.1339
	4	0.8177	0.1004	0.7686	0.1131
	5	0.9011	0.0739	0.8738	0.0835
	6	0.9331	0.0608	0.9128	0.0694
	7	0.9619	0.0458	0.9492	0.0530

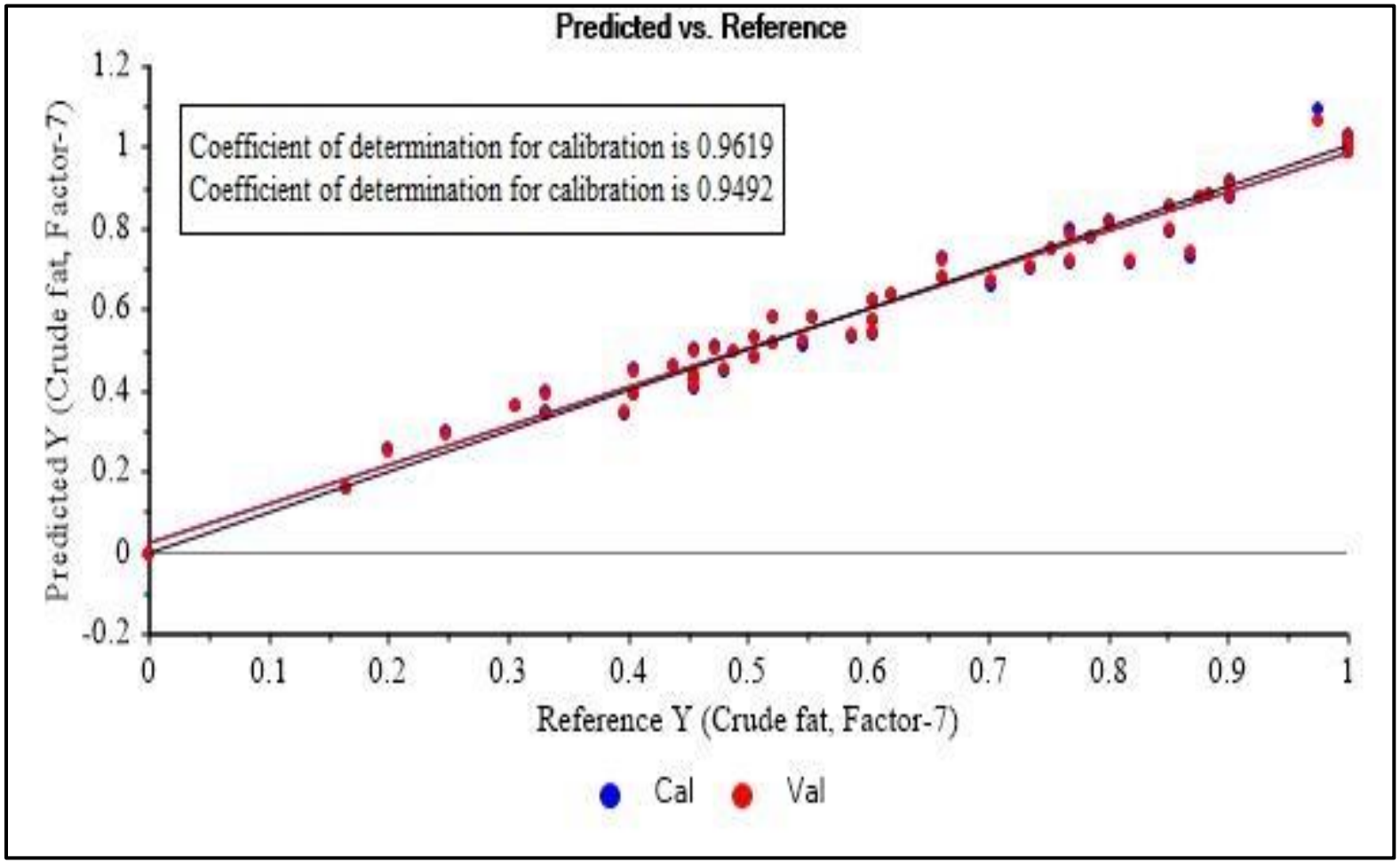


Figure 6. 5 Scatter plot of mango varieties for crude fat content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation with Partial Least Square regression

Table 6. 10 interval Partial Least Square regression model evaluation of crude fat within the range of 700-1300 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.2667	0.2015	0.1907	0.2117
3	0.3656	0.1874	0.2600	0.2024
4	0.4417	0.1758	0.3296	0.1926
5	0.5915	0.1504	0.4961	0.1670
6	0.7598	0.1153	0.6868	0.1317
7	0.8516	0.0906	0.8050	0.1039

Table 6. 11 interval Partial Least Square regression model evaluation of crude fat within the range of 1300-1900 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.3346	0.1919	0.2172	0.2082
3	0.6498	0.1392	0.5945	0.1498
4	0.7947	0.1066	0.7438	0.1191
5	0.8990	0.0747	0.8673	0.0857
6	0.9367	0.0591	0.9131	0.0693
7	0.9624	0.0456	0.9468	0.0542

Table 6. 12 interval Partial Least Square regression model evaluation of crude fat within the range of 1900-2500 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.4764	0.1703	0.4266	0.1782
3	0.6402	0.1411	0.5828	0.1520
4	0.7666	0.1136	0.7026	0.1283
5	0.8633	0.0870	0.8331	0.0961
6	0.8976	0.0752	0.8690	0.0851
7	0.9395	0.0578	0.9153	0.0684

Table 6. 13 interval Partial Least Square regression model evaluation of crude fat with the interval of 300 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-1000	0.5863	0.1513	0.4166	0.1794
1000-1300	0.8440	0.0929	0.7815	0.1100
1300-1600	0.8715	0.0843	0.8145	0.1013
1600-1900	0.8948	0.0763	0.8529	0.0902
1900-2200	0.9284	0.0629	0.9046	0.0726
2200-2500	0.8302	0.0969	0.7676	0.1134

Table 6. 14 interval Partial Least Square regression model evaluation of crude fat with the interval of 200 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-900	0.6474	0.1397	0.5241	0.1623
900-1100	0.5588	0.1563	0.3887	0.1840
1100-1300	0.8900	0.0780	0.8473	0.0919
1300-1500	0.7806	0.1102	0.6772	0.1337
1500-1700	0.8879	0.0787	0.8480	0.0917
1700-1900	0.8490	0.0914	0.7928	0.1071
1900-2100	0.8524	0.0904	0.7898	0.1078
2100-2300	0.8877	0.0788	0.8563	0.0891
2300-2500	0.7145	0.1257	0.6160	0.1458

Table 6. 15 synergic interval Partial Least Square regression model evaluation of crude fat with some selected wavelengths

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
1100-1300, 1500-2300	0.9961	0.0032	0.9044	0.0021

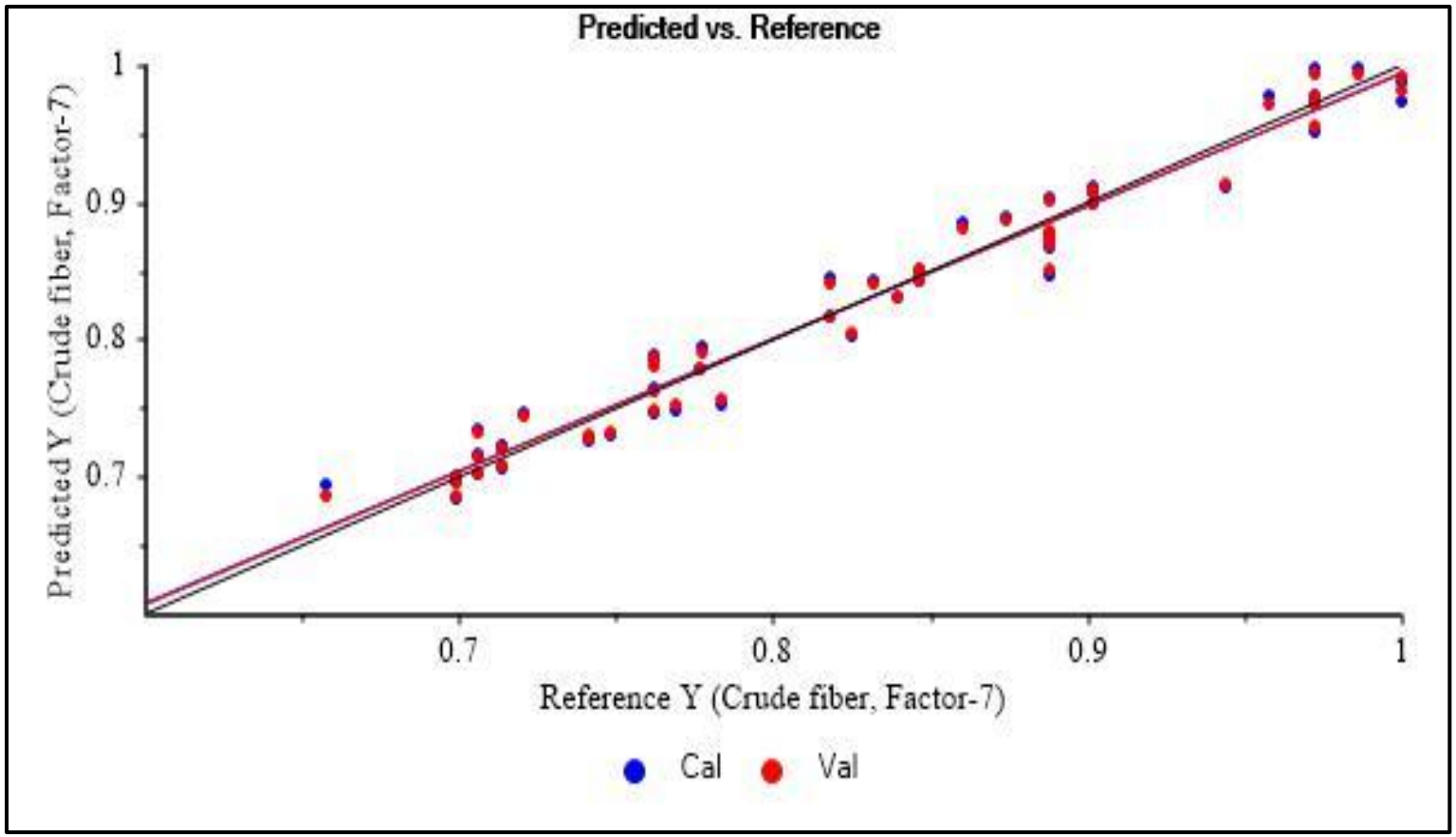


Figure 6. 6 Scatter plot of mango varieties for crude fat content synergic interval Partial Least Square regression in selected wavelength range for calibration and validation

Table 6. 16 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for crude fat with Principal Component Regression method

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.0818	0.2211	NA	0.2365
	3	0.0936	0.2196	NA	0.2383
	4	0.0982	0.2191	NA	0.2424
	5	0.0990	0.2190	NA	0.2483
	6	0.1069	0.2180	NA	0.2521
	7	0.1890	0.2078	NA	0.2406
MSC	2	0.1584	0.1283	0.0546	0.1360
	3	0.1674	0.1276	0.0683	0.1350
	4	0.1700	0.1274	NA	0.1435
	5	0.2508	0.1211	0.0793	0.1342
	6	0.2695	0.1196	NA	0.1433
	7	0.2829	0.1185	NA	0.1721
2^{nd} order derivative	2	0.1717	0.1290	0.0846	0.1357
	3	0.1736	0.1289	0.0194	0.1404
	4	0.2217	0.1251	0.061	0.1374
	5	0.2936	0.1192	0.1134	0.1335
	6	0.2943	0.1191	0.0764	0.1363
	7	0.3228	0.1167	NA	0.1507
MSC + 2^{nd} order derivative	2	0.1586	0.1301	0.0368	0.192
	3	0.1738	0.1289	0.0274	0.1398
	4	0.1769	0.1286	NA	0.1420
	5	0.1857	0.1279	NA	0.1446
	6	0.2261	0.1247	NA	0.1440
	7	0.2267	0.1249	NA	0.1489

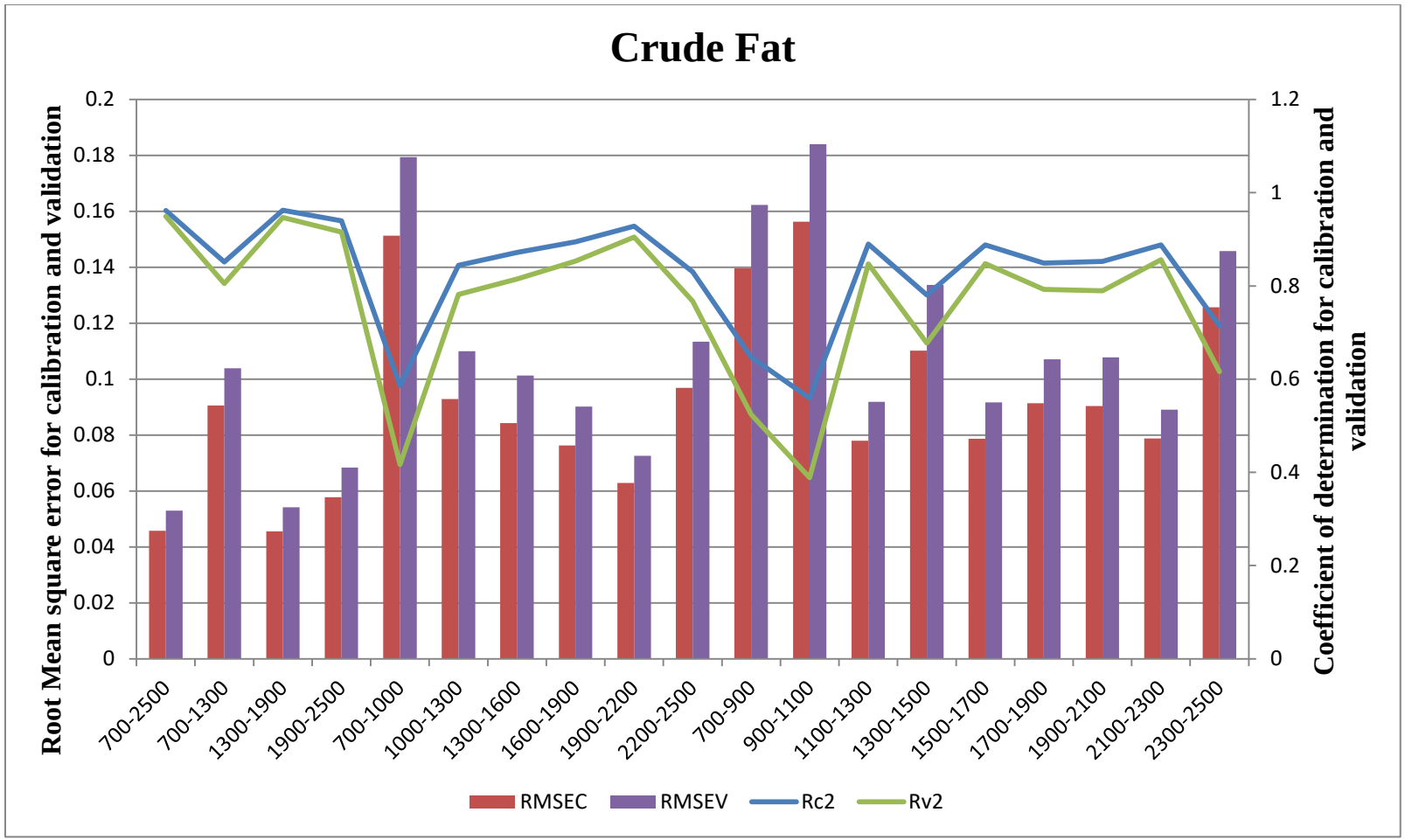


Figure 6. 7 Variation of regression coefficient and root mean square error for calibration and validation with different wavelength range

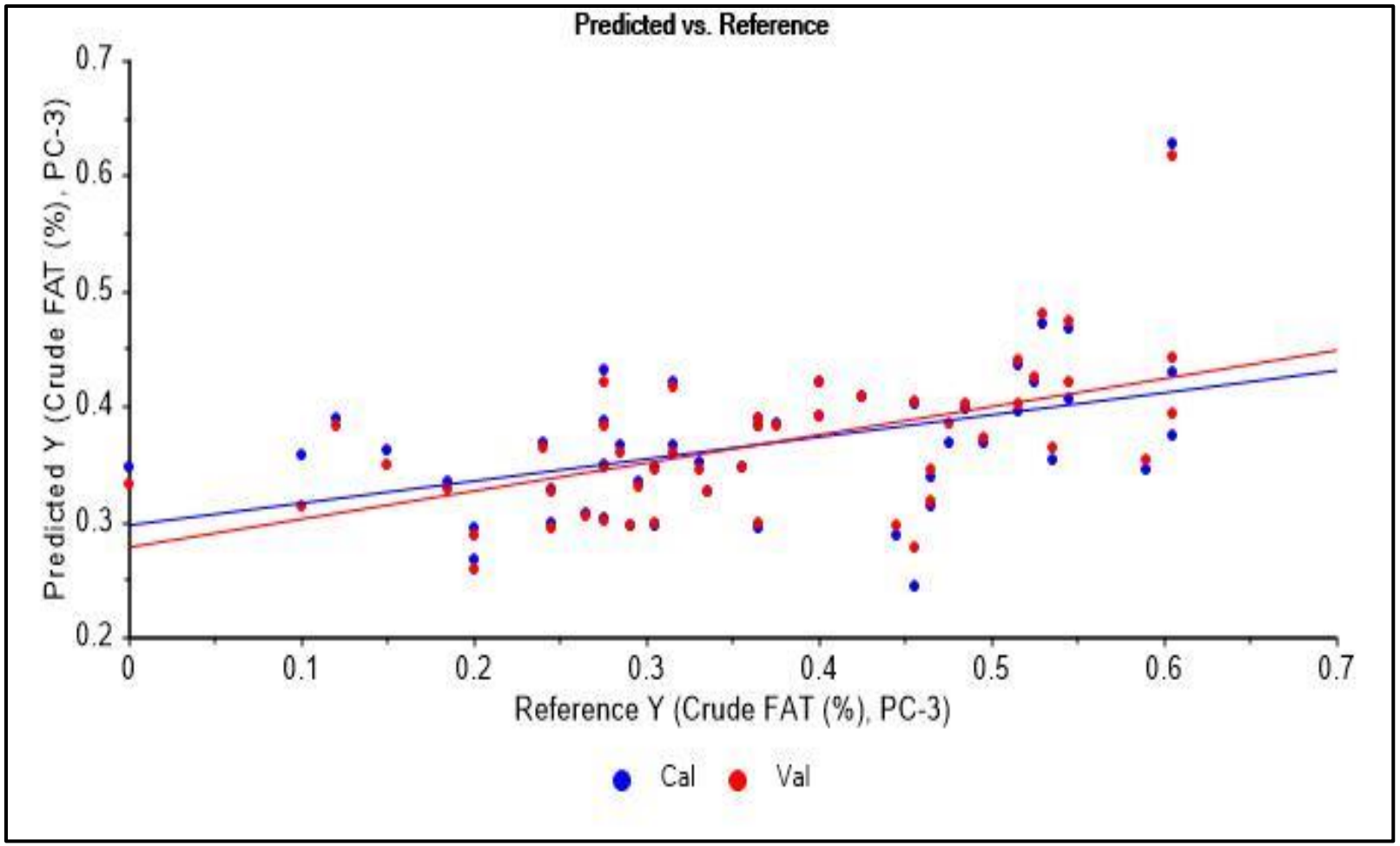


Figure 6. 8 Scatter plot of mango varieties for crude fat content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation with Principal Component Regression

Table 6. 17 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for crude protein with Partial Least Square regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.0937	0.1044	NA	0.1111
	3	0.0978	0.1042	NA	0.1137
	4	0.1091	0.1035	NA	0.1152
	5	0.1493	0.1012	NA	0.1152
	6	0.2017	0.0980	NA	0.1141
	7	0.2937	0.0922	NA	0.1113
MSC	2	0.0705	0.1057	NA	0.1121
	3	0.1228	0.1027	NA	0.1108
	4	0.1536	0.1009	NA	0.1113
	5	0.1997	0.0981	NA	0.1097
	6	0.2873	0.0926	0.0517	0.1068
	7	0.4754	0.0794	0.2397	0.0956
2^{nd} order derivative	2	0.6346	0.0663	0.5876	0.0704
	3	0.7601	0.0537	0.7203	0.0580
	4	0.8729	0.0391	0.8480	0.0427
	5	0.9505	0.0243	0.9381	0.0272
	6	0.9732	0.0179	0.9647	0.0206
	7	0.9816	0.0148	0.9740	0.0176
MSC + 2^{nd} order derivative	2	0.6282	0.0669	0.5799	0.0711
	3	0.7619	0.0535	0.7228	0.0577
	4	0.8727	0.0391	0.8471	0.0429
	5	0.9538	0.0235	0.9426	0.0262
	6	0.9726	0.0181	0.9642	0.0207
	7	0.9822	0.0146	0.9744	0.0175

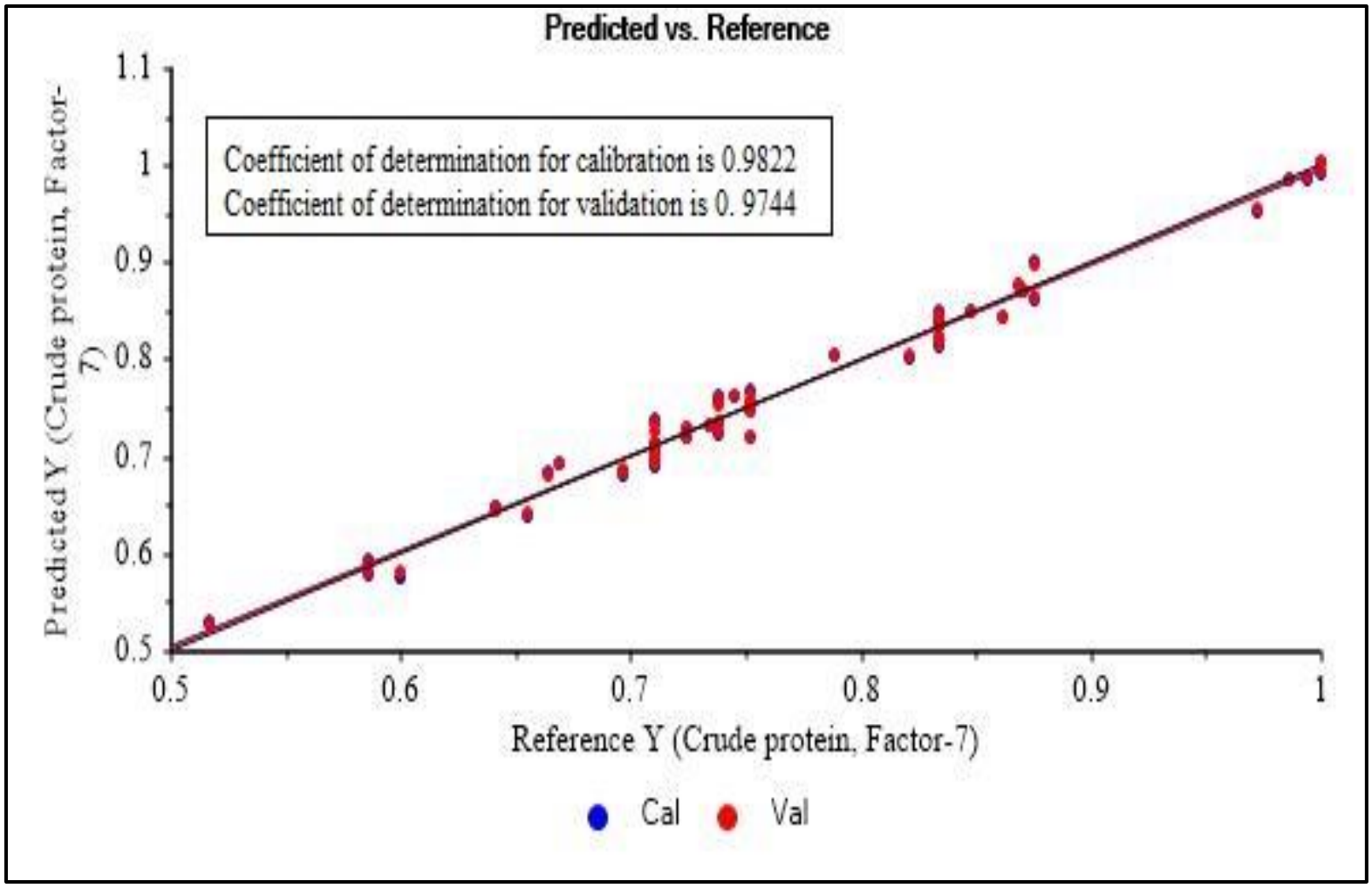


Figure 6. 9 Scatter plot of mango varieties for crude protein content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation

Table 6. 18 interval Partial Least Square regression model evaluation of crude protein within the range of 700-1300 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.1700	0.0999	0.0696	0.1058
3	0.2118	0.0974	0.0806	0.1052
4	0.2772	0.0935	0.1282	0.1024
5	0.3786	0.0864	0.2240	0.0966
6	0.6092	0.0685	0.3881	0.0858
7	0.6602	0.0639	0.4665	0.0801

Table 6. 19 interval Partial Least Square regression model evaluation of crude protein within the range of 1300-1900 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.6186	0.0677	0.5650	0.0723
3	0.8004	0.0490	0.7643	0.0532
4	0.8593	0.0411	0.8338	0.0447
5	0.8909	0.0362	0.8653	0.0402
6	0.9574	0.0226	0.9443	0.0253
7	0.9790	0.0158	0.9704	0.0188

Table 6. 20 interval Partial Least Square regression model evaluation of crude protein within the range of 1900-2500 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.6386	0.0659	0.6013	0.0692
3	0.6981	0.0602	0.6591	0.0640
4	0.7947	0.0497	0.7567	0.0541
5	0.8734	0.0390	0.8477	0.0428
6	0.9140	0.0321	0.8886	0.0366

7	0.9392	0.0270	0.9187	0.0312

Table 6. 21 interval Partial Least Square regression model evaluation of crude protein with the interval of 300 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-1000	0.3796	0.0864	0.1312	0.1022
1000-1300	0.6882	0.0612	0.5156	0.0763
1300-1600	0.8555	0.0416	0.7884	0.0504
1600-1900	0.8384	0.0441	0.7826	0.0511
1900-2200	0.9244	0.0301	0.8823	0.0376
2200-2500	0.8675	0.0399	0.8308	0.0451

Table 6. 22 interval Partial Least Square regression model evaluation of crude protein with the interval of 200 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-900	0.3584	0.0878	0.1500	0.1011
900-1100	0.3367	0.0893	NA	0.1101
1100-1300	0.7347	0.0565	0.6152	0.0680
1300-1500	0.6380	0.0198	0.6726	0.0241
1500-1700	0.8792	0.0381	0.8157	0.0470
1700-1900	0.7462	0.0552	0.6674	0.0632
1900-2100	0.8206	0.0464	0.7470	0.0551
2100-2300	0.7865	0.0506	0.7035	0.0597
2300-2500	0.8031	0.0486	0.7476	0.0551

Table 6. 23 synergic interval Partial Least Square regression model evaluation of crude protein with some selected wavelengths

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
1900-2100, 2300-2500	0.9901	0.0088	0.9155	0.0197

1500-1700, 1900-2100, 2300-2500	0.9343	0.0468	0.9231	0.0561

Table 6. 24 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for crude protein with Principal Component Regression

Preprocessing techniques	**Factors**	**R_c^2**	**RMSEC**	**R_v^2**	**RMSEV**
Original spectra	2	0.06606	0.1060	NA	0.1122
	3	0.0919	0.1045	NA	0.1142
	4	0.0946	0.1044	NA	0.1164
	5	0.0992	0.1042	NA	0.1197
	6	0.0999	0.1041	NA	0.1211
	7	0.1260	0.1025	NA	0.1207
MSC	2	0.1430	0.6994	0.0388	0.7407
	3	0.1560	0.6941	NA	0.7566
	4	0.1680	0.6892	NA	0.7836
	5	0.1697	0.6885	NA	0.8062
	6	0.2051	0.6736	NA	0.9089
	7	0.2424	0.6576	NA	0.8043
2^{nd} order derivative	2	0.1675	0.6928	0.0921	0.7227
	3	0.1689	0.6914	NA	0.8416
	4	0.1749	0.6889	NA	0.8991
	5	0.2285	0.6662	NA	0.8950
	6	0.2608	0.6521	NA	0.8594
	7	0.2810	0.6431	NA	0.8225
MSC + 2^{nd} order derivative	2	0.1271	0.7086	0.0228	0.7498
	3	0.1545	0.6974	0.0146	0.7529
	4	0.2657	0.6499	0.1190	0.7119
	5	0.2864	0.6407	0.1069	0.7168
	6	0.3017	0.6338	0.0611	0.7349
	7	0.3020	0.6333	0.0148	0.7528

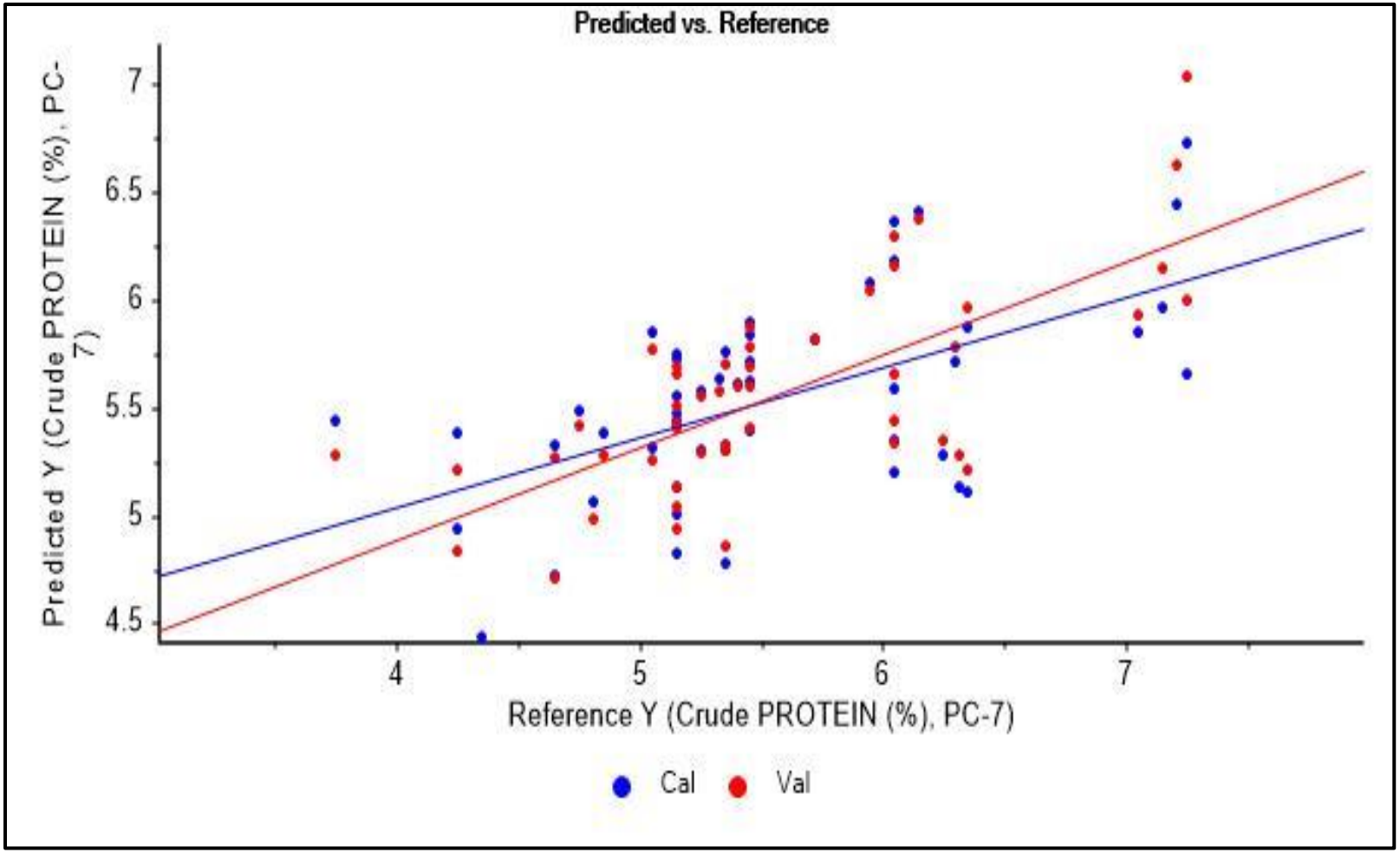


Figure 6. 10 Scatter plot of mango varieties for crude protein content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation with Principal Component Regression

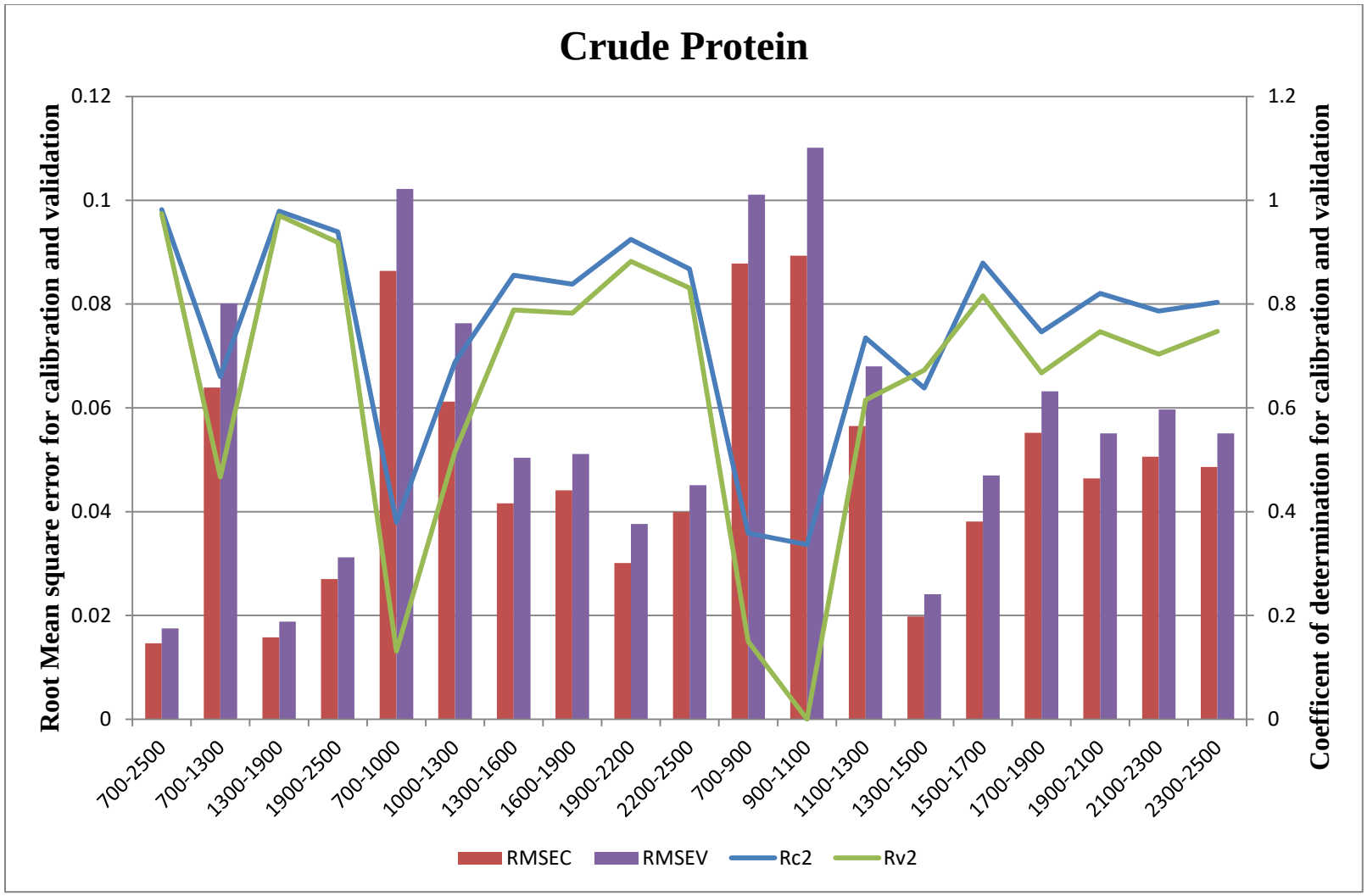


Figure 6. 11 Variation of regression coefficient and root mean square error for calibration and validation with different wavelength range

Table 6. 25 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for crude fiber with Partial Least Square regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.1807	0.0871	0.0984	0.0914
	3	0.2142	0.0853	0.1063	0.0910
	4	0.2211	0.0849	0.0767	0.0923
	5	0.2599	0.0828	0.1035	0.0911
	6	0.4356	0.0723	0.2691	0.0823
	7	0.4997	0.0681	0.2904	0.0811
MSC	2	0.1475	0.0889	0.0570	0.0935
	3	0.1968	0.0863	0.0877	0.0919
	4	0.2306	0.0844	0.0896	0.0918
	5	0.3696	0.0764	0.2156	0.0852
	6	0.4730	0.0699	0.2933	0.0809
	7	0.5679	0.0633	0.3648	0.0767
2^{nd} order derivative	2	0.6031	0.0606	0.5594	0.0639
	3	0.7508	0.0480	0.7112	0.0517
	4	0.8452	0.0378	0.8118	0.0417
	5	0.9137	0.0282	0.8945	0.0312
	6	0.9436	0.0228	0.9279	0.0258
	7	0.9709	0.0164	0.9582	0.0196
MSC + 2^{nd} order derivative	2	0.5900	0.0616	0.5461	0.0648
	3	0.7460	0.0485	0.7065	0.0521
	4	0.8461	0.0377	0.8128	0.0416
	5	0.9177	0.0276	0.8991	0.0305
	6	0.9461	0.0223	0.9308	0.0253
	7	0.9712	0.0163	0.9590	0.0194

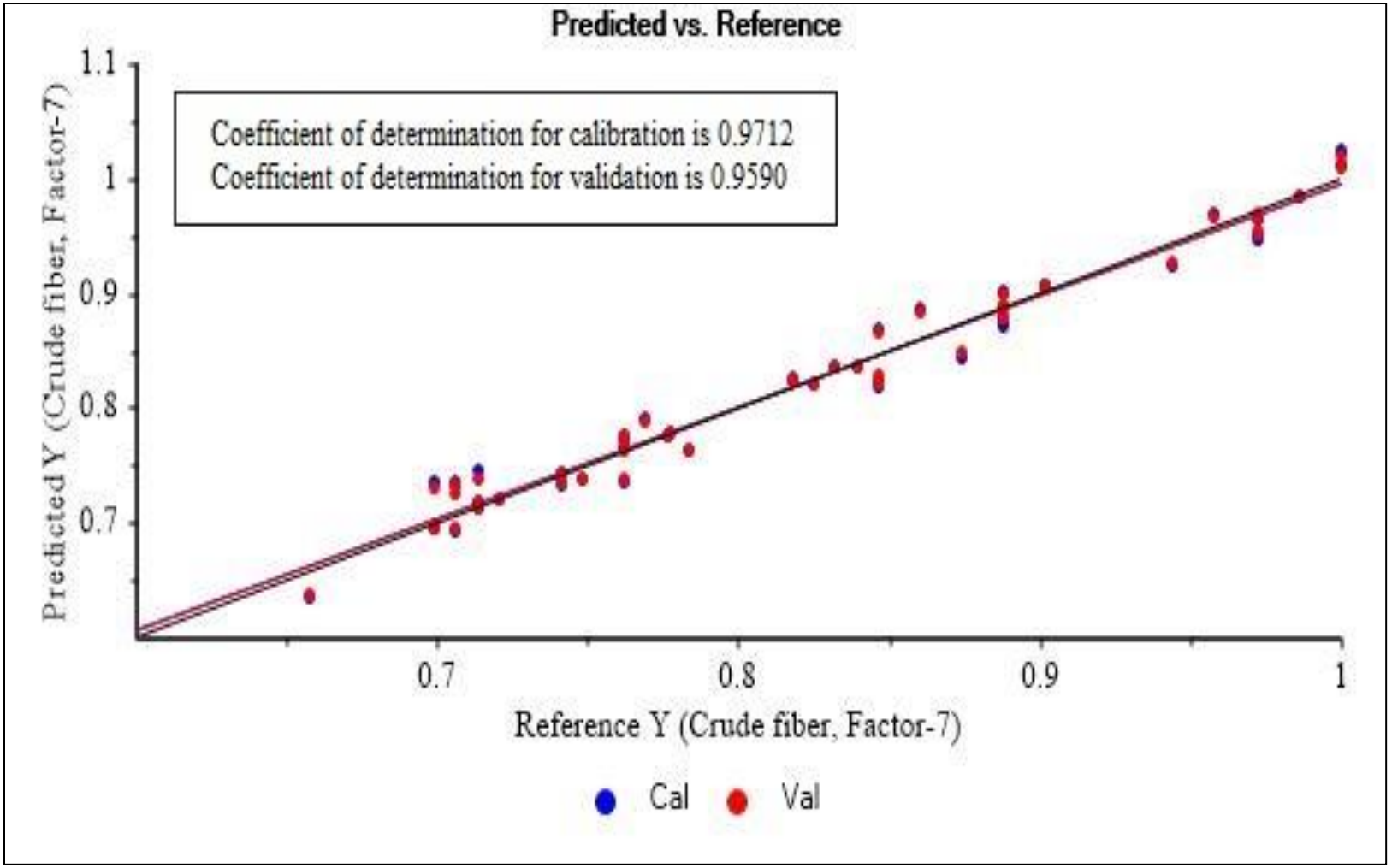


Figure 6. 12 Scatter plot of mango varieties for crude fiber content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation

Table 6. 26 interval Partial Least Square regression model evaluation of crude fiber within the range of 700-1300 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.2065	0.0857	0.1377	0.0894
3	0.3379	0.0783	0.2420	0.0838
4	0.5410	0.0652	0.4511	0.0713
5	0.6060	0.0604	0.5047	0.0677
6	0.6966	0.0530	0.5390	0.0614
7	0.7529	0.0478	0.6579	0.0563

Table 6. 27 interval Partial Least Square regression model evaluation of crude fiber within the range of 1300-1900 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.5820	0.0622	0.5314	0.0659
3	0.7469	0.0484	0.7004	0.0527
4	0.8047	0.0425	0.7645	0.0467
5	0.9155	0.0279	0.8898	0.0319
6	0.9561	0.0201	0.9392	0.0237
7	0.9765	0.0147	0.9642	0.0182

Table 6. 28 interval Partial Least Square regression model evaluation of crude fiber within the range of 1900-2500 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.5157	0.0670	0.4723	0.0699
3	0.6365	0.0580	0.6015	0.0607
4	0.7839	0.0447	0.7322	0.0498
5	0.8560	0.0365	0.8181	0.0410
6	0.8979	0.0307	0.8648	0.0354
7	0.9274	0.0259	0.9037	0.0298

Table 6. 29 interval Partial Least Square regression model evaluation of crude fiber with the interval of 300 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-1000	0.6575	0.0563	0.5229	0.0665
1000-1300	0.7880	0.0443	0.7043	0.0523
1300-1600	0.8969	0.0309	0.8590	0.0361
1600-1900	0.9260	0.0261	0.8996	0.0305
1900-2200	0.9511	0.0212	0.9342	0.0247
2200-2500	0.8577	0.0363	0.8194	0.0409

Table 6. 30 interval Partial Least Square regression model evaluation of crude fiber with the interval of 200 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-900	0.7216	0.0508	0.6120	0.0599
900-1100	0.6163	0.0596	0.4437	0.0718
1100-1300	0.7909	0.0440	0.7245	0.0505
1300-1500	0.7637	0.0468	0.6846	0.0540
1500-1700	0.8898	0.0319	0.8580	0.0362
1700-1900	0.8541	0.0367	0.7977	0.0433
1900-2100	0.8568	0.0364	0.8075	0.0422
2100-2300	0.8434	0.0381	0.7898	0.0441
2300-2500	0.7494	0.0482	0.6621	0.0559

Table 6. 31 synergic interval Partial Least Square regression model evaluation of crude fiber with some selected wavelengths

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
1500-1700, 1900-2100	0.9563	0.0081	0.9559	0.0154
1900-2100, 2300-2500	0.9752	0.0065	0.9607	0.0149

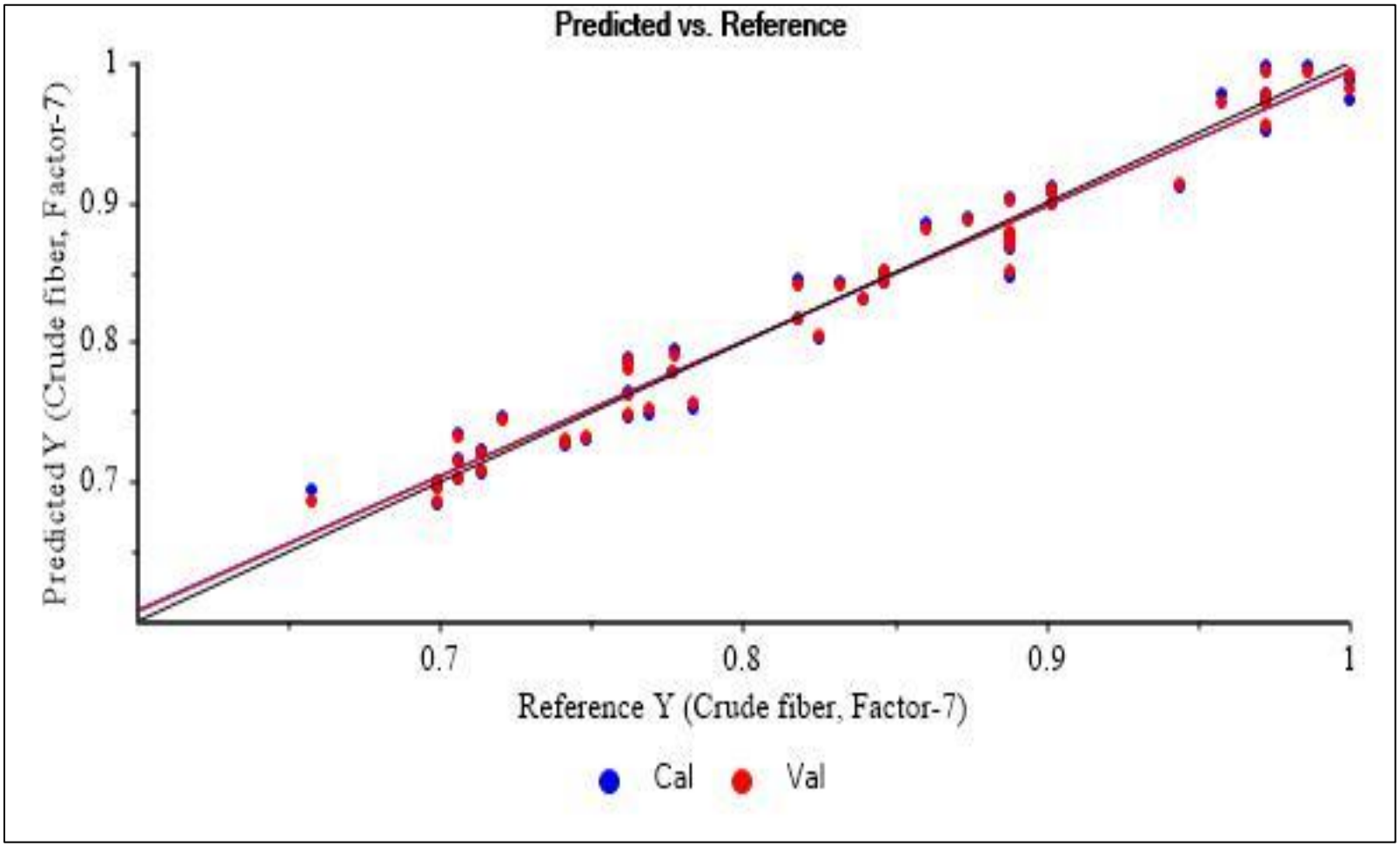


Figure 6. 13 Scatter plot of mango varieties for crude fiber content synergic interval Partial Least Square regression in selected wavelength range for calibration and validation

Table 6. 32 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for crude fiber with Principal Component Regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.0804	0.0923	NA	0.0979
	3	0.1402	0.0893	NA	0.0965
	4	0.1937	0.0864	0.0291	0.0948
	5	0.2433	0.0837	0.0819	0.0922
	6	0.2439	0.0832	0.0424	0.0942
	7	0.2455	0.0822	0.0063	0.0960
MSC	2	0.1656	0.0642	0.0159	0.0698
	3	0.1934	0.062	NA	0.0711
	4	0.1952	0.0631	NA	0.0749
	5	0.2037	0.0628	NA	0.0792
	6	0.2351	0.0615	NA	0.1016
	7	0.2472	0.0610	NA	0.1074
2^{nd} order derivative	2	0.1298	0.0659	NA	0.0734
	3	0.1638	0.0646	NA	0.0774
	4	0.1660	0.0645	NA	0.0819
	5	0.1667	0.0642	NA	0.0834
	6	0.2179	0.0625	NA	0.0794
	7	0.2221	0.0623	NA	0.0942
MSC + 2^{nd} order derivative	2	0.0943	0.0672	NA	0.0709
	3	0.1227	0.0660	NA	0.0710
	4	0.1542	0.0649	NA	0.0717
	5	0.1951	0.0634	NA	0.0718
	6	0.2619	0.0607	0.0084	0.0703
	7	0.2679	0.0604	NA	0.0719

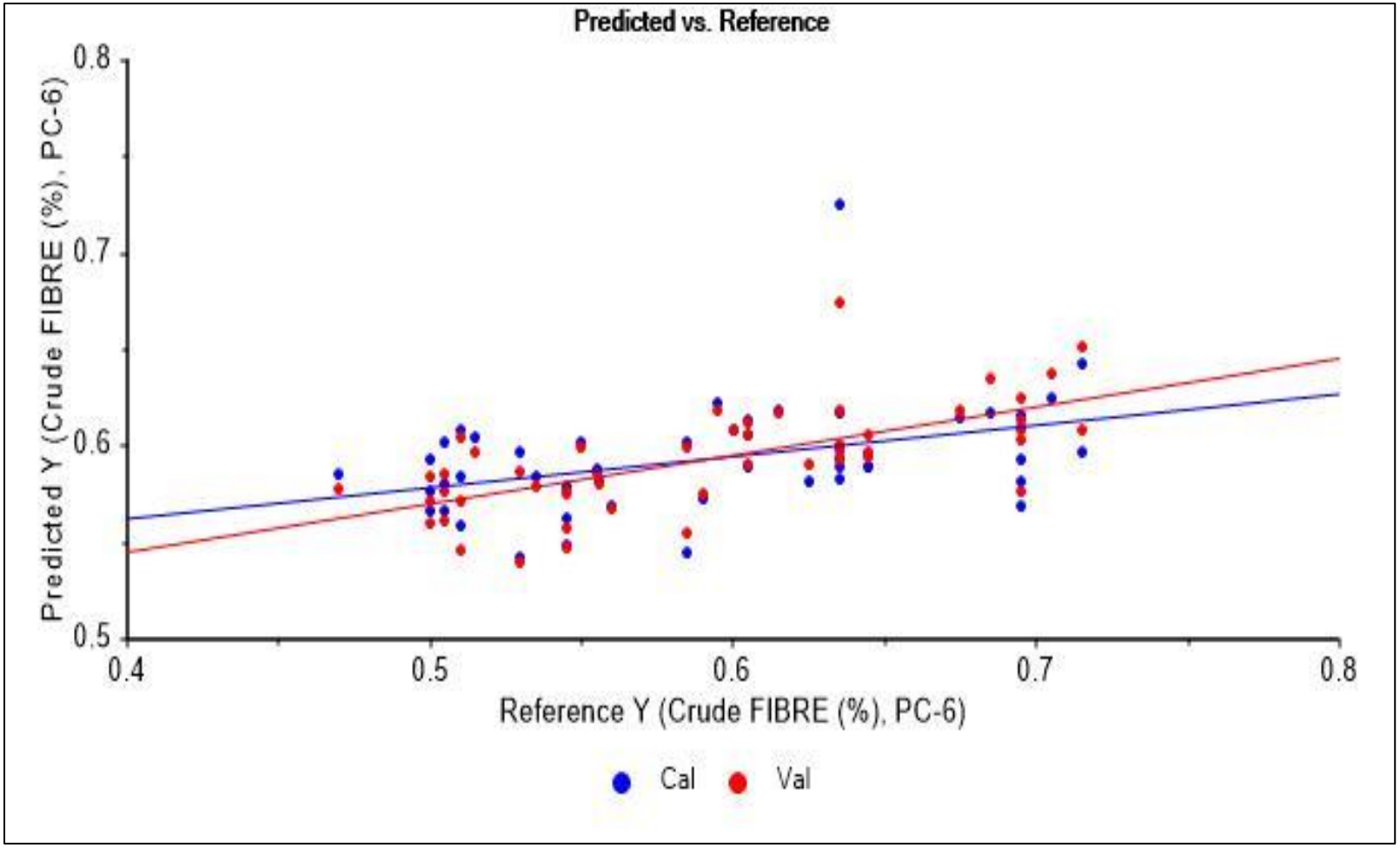


Figure 6. 14 Scatter plot of mango varieties for crude fiber content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation with Principal Component Regression

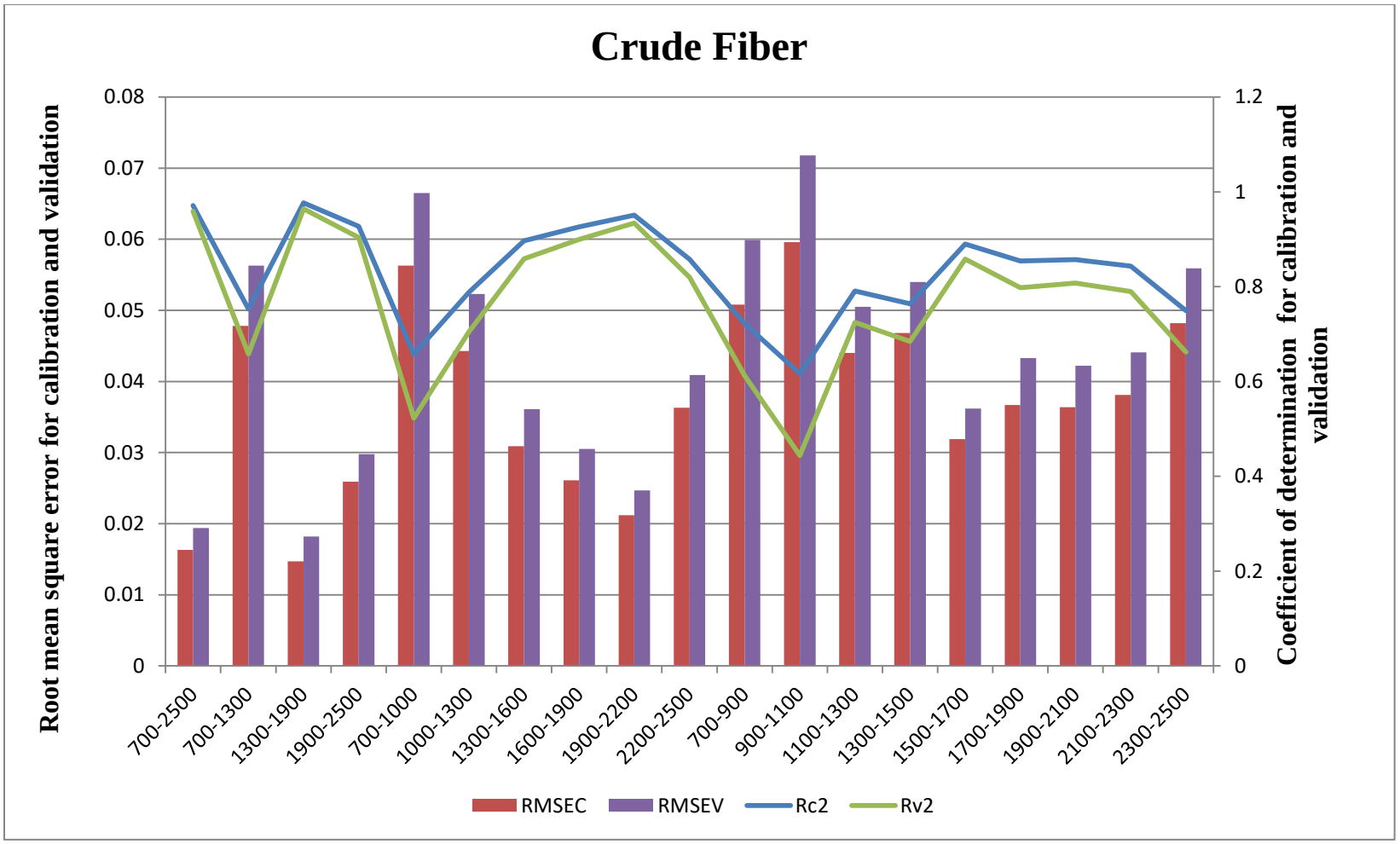


Figure 6. 15 Variation of regression coefficient and root mean square error for calibration and validation with different wavelength range

Table 6. 33 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for pH with Partial Least Square regression

Preprocessing techniques	**Factors**	**R_c^2**	**RMSEC**	**R_v^2**	**RMSEV**
Original spectra	2	0.0826	0.0879	NA	0.0938
	3	0.1140	0.0864	NA	0.0943
	4	0.1291	0.0857	NA	0.0947
	5	0.1987	0.0822	NA	0.0924
	6	0.2892	0.0774	0.0216	0.0908
	7	0.4779	0.0663	0.2955	0.0771
MSC	2	0.0937	0.0874	NA	0.0938
	3	0.1859	0.0828	0.0463	0.0897
	4	0.2341	0.0803	0.0168	0.0910
	5	0.4177	0.0700	0.2608	0.0789
	6	0.5228	0.0634	0.3549	0.0737
	7	0.6039	0.0578	0.3935	0.0715
2^{nd} order derivative	2	0.6597	0.0535	0.6179	0.0567
	3	0.8470	0.0359	0.8228	0.0386
	4	0.9182	0.0262	0.8984	0.0292
	5	0.9683	0.0163	0.9552	0.0194
	6	0.9840	0.0116	0.9764	0.0140
	7	0.9905	0.0089	0.9856	0.0010
MSC + 2^{nd} order derivative	2	0.6590	0.0536	0.6176	0.0567
	3	0.8451	0.0361	0.8205	0.0389
	4	0.9178	0.0263	0.8976	0.0293
	5	0.9675	0.0165	0.9550	0.0194
	6	0.9836	0.0117	0.9768	0.0139
	7	0.9913	0.0085	0.9868	0.0105

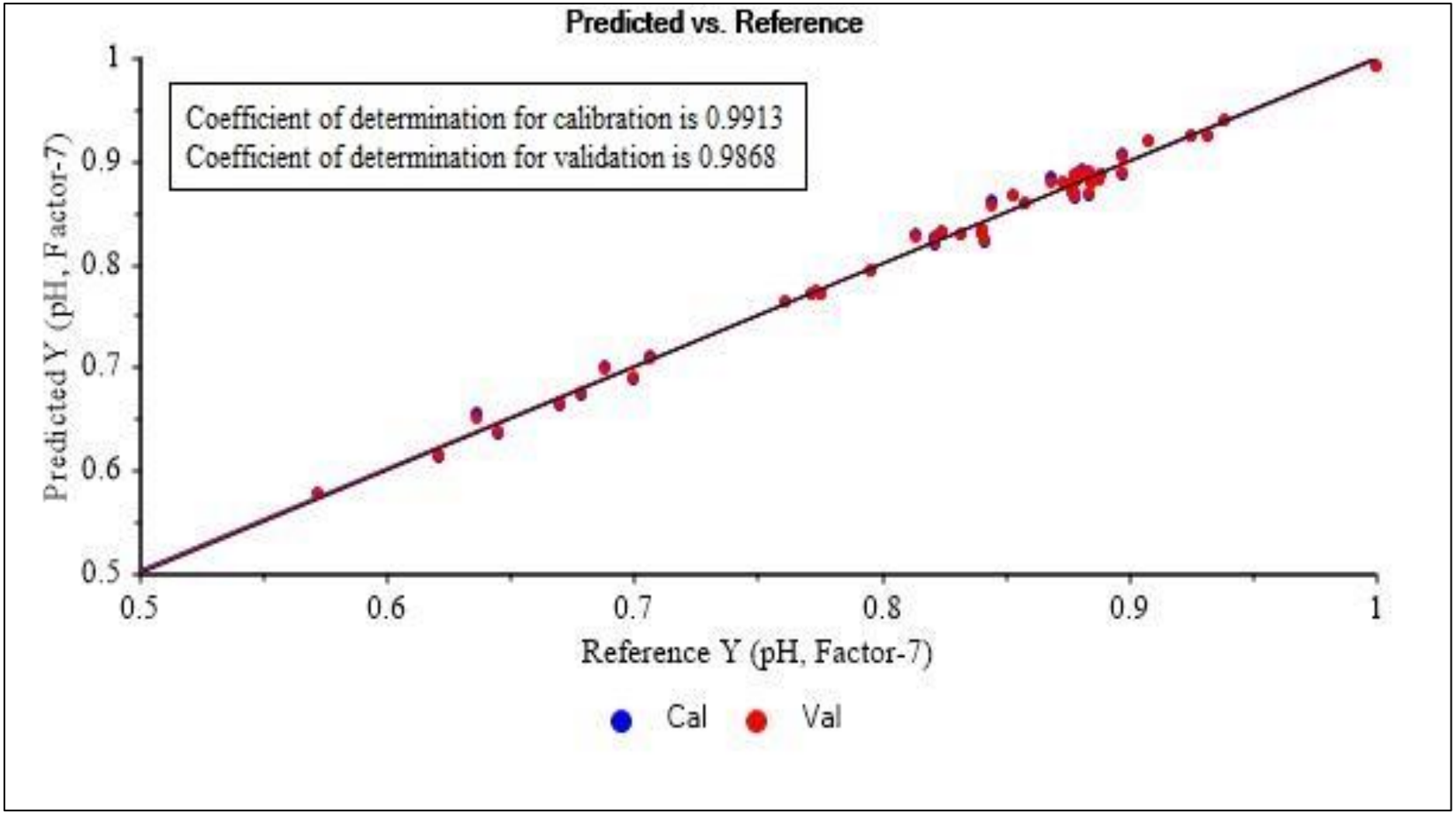


Figure 6. 16 Scatter plot of mango varieties for pH content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation with Partial Least Square regression

Table 6. 34 interval Partial Least Square regression model evaluation of pH within the range of 700-1300 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.2112	0.0815	0.0648	0.0888
3	0.2934	0.0777	0.0576	0.0891
4	0.3521	0.0739	0.1065	0.0868
5	0.5096	0.0643	0.2805	0.0779
6	0.7448	0.0463	0.6590	0.0536
7	0.7819	0.0428	0.6918	0.0509

Table 6. 35 interval Partial Least Square regression model evaluation of pH within the range of 1300-1900 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.4090	0.0706	0.3351	0.0749
3	0.6385	0.0552	0.5758	0.0598
4	0.8029	0.0407	0.7584	0.0451
5	0.9112	0.0273	0.8853	0.0311
6	0.9434	0.0218	0.9241	0.0252
7	0.9700	0.0158	0.9554	0.0193

Table 6. 36 interval Partial Least Square regression model evaluation of pH within the range of 1900-2500 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.6676	0.0529	0.5951	0.0584
3	0.7749	0.0435	0.7208	0.0485
4	0.8824	0.0314	0.8497	0.0356
5	0.9365	0.0231	0.9171	0.0264
6	0.9584	0.0187	0.9410	0.0223
7	0.9761	0.0141	0.9668	0.0167

Table 6. 37 interval Partial Least Square regression model evaluation of pH with the interval of 300 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-1000	0.6786	0.0520	0.5536	0.0613
1000-1300	0.7787	0.0432	0.6785	0.0520
1300-1600	0.9180	0.0262	0.8789	0.0319
1600-1900	0.9351	0.0233	0.9050	0.0283
1900-2200	0.9489	0.0207	0.9280	0.0246
2200-2500	0.9246	0.0252	0.8887	0.0306

Table 6. 38 interval Partial Least Square regression model evaluation of pH with the interval of 200 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-900	0.6256	0.0562	0.4652	0.0617
900-1100	0.5248	0.0633	0.3430	0.0744
1100-1300	0.7757	0.0435	0.6727	0.0525
1300-1500	0.8452	0.0361	0.7796	0.0431
1500-1700	0.9267	0.0248	0.9041	0.0288
1700-1900	0.8347	0.0373	0.7584	0.0451
1900-2100	0.8363	0.0371	0.7940	0.0416
2100-2300	0.9092	0.0276	0.8703	0.0330
2300-2500	0.8174	0.0392	0.7382	0.0470

Table 6. 39 siPLS Regression model evaluation of pH with some selected wavelengths

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
1500-1700, 2100-2300	0.9800	0.0164	0.9553	0.0196
1500-1700, 1900-2100	0.9532	0.0466	0.9399	0.0521
1500-1700, 1900-2300	0.9552	0.0270	0.9378	0.0341

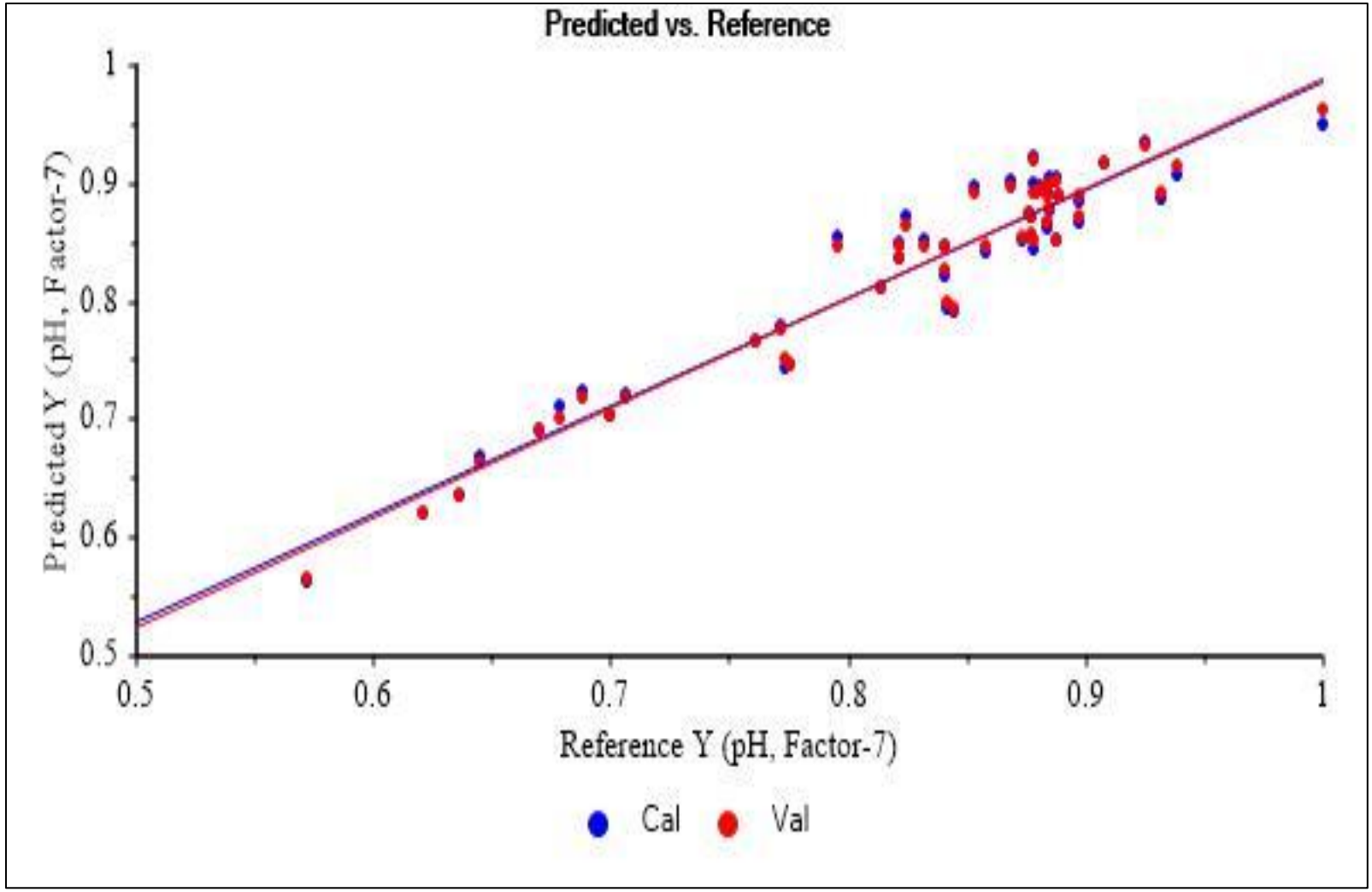


Figure 6. 17 Scatter plot of mango varieties for pH content synergic interval Partial Least Square regression in selected wavelength range for calibration and validation

Table 6. 40 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for pH with Principal Component Regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.0428	0.0898	NA	0.0963
	3	0.0555	0.0892	NA	0.0967
	4	0.0940	0.0874	NA	0.0973
	5	0.1299	0.0856	NA	0.0975
	6	0.1608	0.0841	NA	0.0992
	7	0.1697	0.0837	NA	0.1021
MSC	2	0.0703	0.4299	NA	0.4630
	3	0.0781	0.4281	NA	0.4993
	4	0.0783	0.4280	NA	0.5668
	5	0.0793	0.4278	NA	0.5736
	6	0.1034	0.4222	NA	0.7190
	7	0.1302	0.4158	NA	0.7407
2^{nd} order derivative	2	0.0704	0.4265	NA	0.4671
	3	0.0888	0.4222	NA	0.4955
	4	0.0918	0.4215	NA	0.4978
	5	0.0938	0.4211	NA	0.5045
	6	0.1085	0.4176	NA	0.5012
	7	0.1108	0.4171	NA	0.5196
MSC + 2^{nd} order derivative	2	0.1096	0.4174	NA	0.4480
	3	0.1099	0.4173	NA	0.4534
	4	0.1363	0.4111	NA	0.4526
	5	0.2135	0.3923	NA	0.4443
	6	0.3218	0.3642	0.0749	0.4254
	7	0.3356	0.3605	0.0390	0.4336

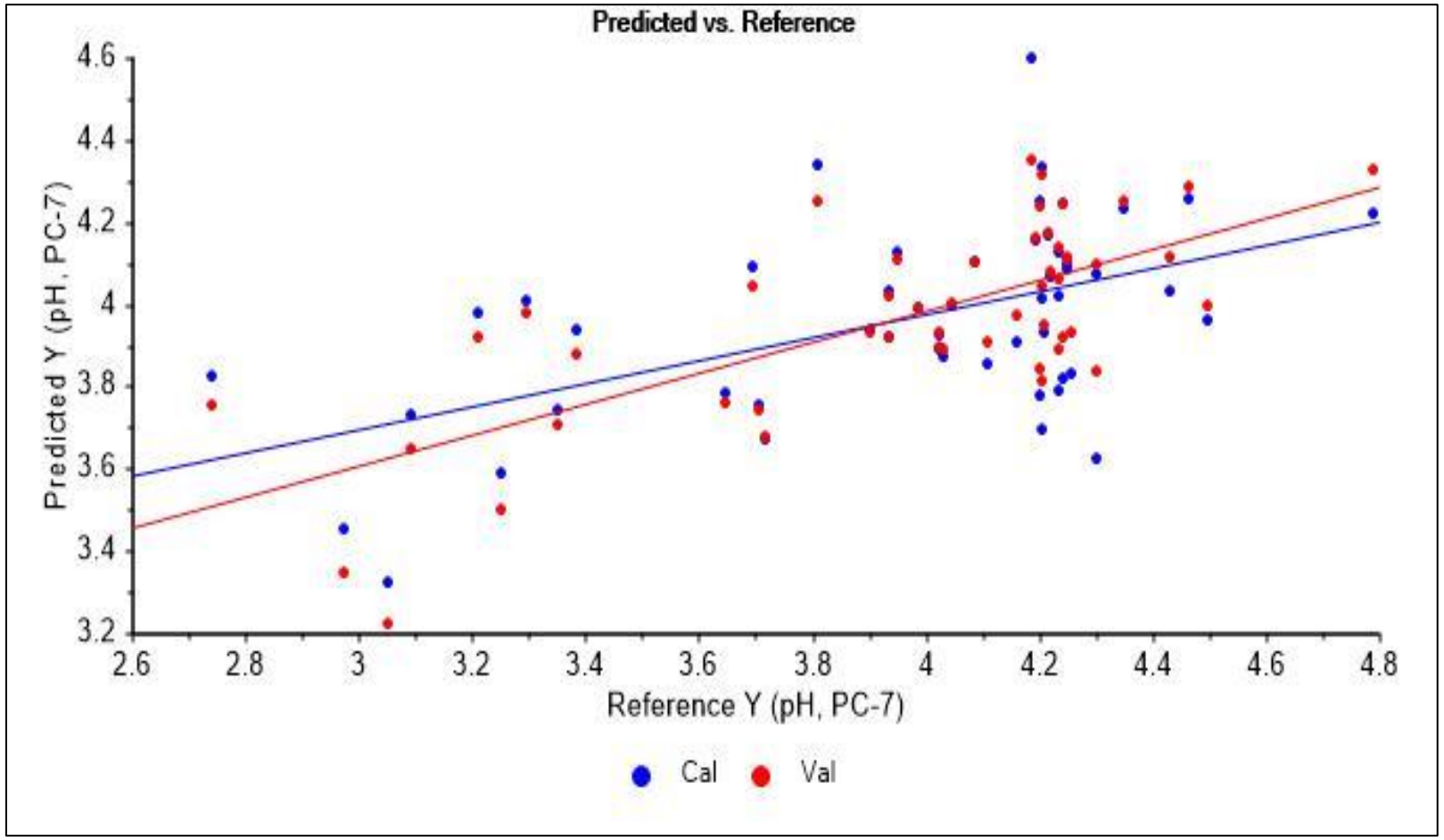


Figure 6. 18 Scatter plot of mango varieties for pH content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation Principal Component Regression

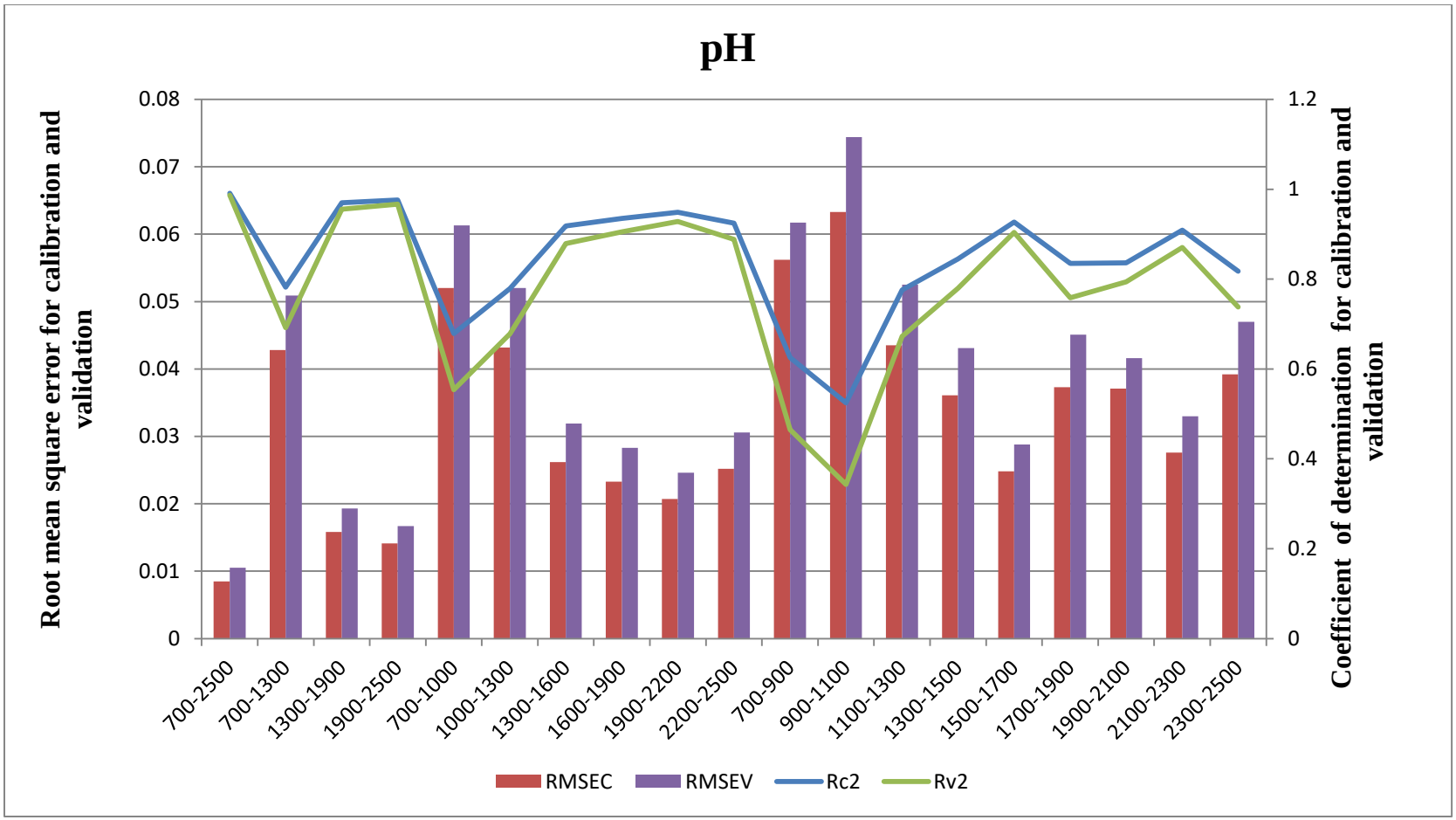


Figure 6. 19 Variation of regression coefficient and root mean square error for calibration and validation with different wavelength range

Table 6. 41 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for carbohydrate with Partial Least Square regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.3121	0.1875	0.2134	0.2006
	3	0.4119	0.1734	0.2734	0.1928
	4	0.4422	0.1689	0.2837	0.1914
	5	0.4817	0.1628	0.2489	0.1960
	6	0.5551	0.1508	0.3447	0.1831
	7	0.5945	0.1440	0.3866	0.1771
MSC	2	0.0934	0.2153	NA	0.2308
	3	0.2540	0.1953	0.1145	0.2128
	4	0.3542	0.1814	0.1745	0.2055
	5	0.4315	0.1705	0.2539	0.1953
	6	0.5488	0.1519	0.3750	0.1788
	7	0.6459	0.1328	0.4922	0.1611
2^{nd} order derivative	2	0.6715	0.1296	0.6342	0.1367
	3	0.8523	0.0869	0.8254	0.0944
	4	0.9112	0.0673	0.8921	0.0742
	5	0.9649	0.0423	0.9576	0.0465
	6	0.9800	0.0319	0.9738	0.0365
	7	0.9910	0.0213	0.9875	0.0252
MSC + 2^{nd} order derivative	2	0.6715	0.1296	0.6328	0.1370
	3	0.8448	0.0891	0.8159	0.0970
	4	0.9088	0.0682	0.8893	0.0752
	5	0.9675	0.0407	0.9604	0.0449
	6	0.9823	0.0300	0.9768	0.0343
	7	0.9923	0.0197	0.9892	0.0234

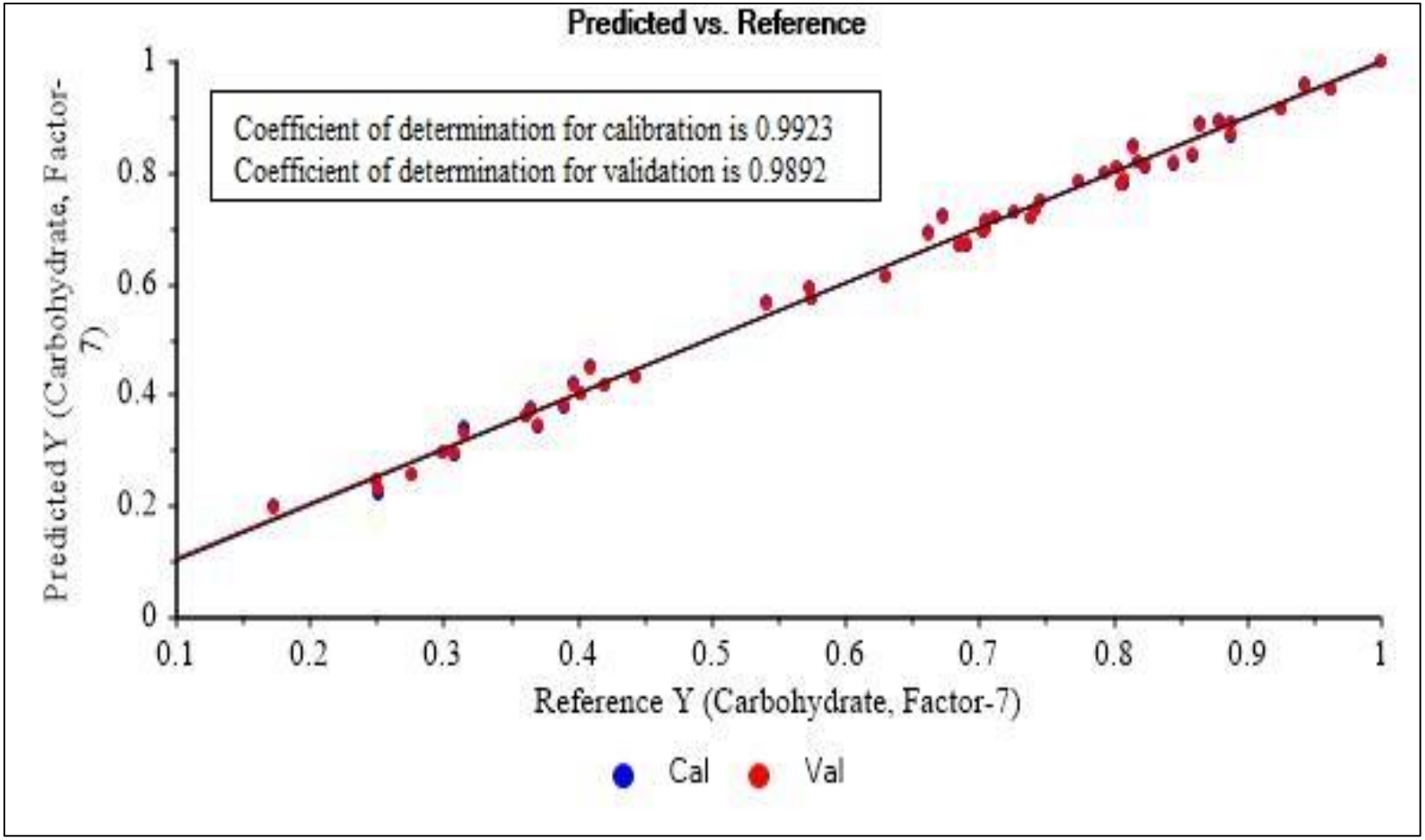


Figure 6. 20 Scatter plot of mango varieties for carbohydrate content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation

Table 6. 42 interval Partial Least Square regression model evaluation of carbohydrate within the range of 700-1300 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.2407	0.1970	0.1334	0.2105
3	0.2622	0.1942	0.1475	0.2088
4	0.2845	0.1913	0.1611	0.2071
5	0.4140	0.1731	0.2975	0.1895
6	0.5099	0.1583	0.3508	0.1822
7	0.6270	0.1381	0.5028	0.1594

Table 6. 43 interval Partial Least Square regression model evaluation of carbohydrate within the range of 1300-1900 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.3448	0.1830	0.2453	0.1964
3	05230	0.1562	0.4348	0.1700
4	0.7301	0.1174	0.6542	0.1330
5	0.8635	0.0835	0.8167	0.0968
6	0.9285	0.0604	0.9026	0.0705
7	0.9582	0.0462	0.9394	0.0556

Table 6. 44 interval Partial Least Square regression model evaluation of carbohydrate within the range of 1900-2500 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.6069	0.1418	0.5386	0.1536
3	0.7189	0.1199	0.6795	0.1280
4	0.8341	0.0921	0.8029	0.1004
5	0.9027	0.0705	0.8769	0.0793
6	0.9452	0.0529	0.9259	0.0613
7	0.9658	0.0418	0.9487	0.0512

Table 6. 45 interval Partial Least Square regression model evaluation of carbohydrate with the interval of 300 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-1000	0.5528	0.1512	0.3573	0.1813
1000-1300	0.6452	0.1347	0.5148	0.1575
1300-1600	0.8417	0.0899	0.7822	0.1055
1600-1900	0.8389	0.0907	0.7637	0.1099
1900-2200	0.8682	0.0821	0.8298	0.0932
2200-2500	0.9416	0.0546	0.9175	0.0649

Table 6. 46 interval Partial Least Square regression model evaluation of carbohydrate with the interval of 200 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-900	0.5186	0.1569	0.3478	0.1826
900-1100	0.3755	0.1787	0.1451	0.2091
1100-1300	0.6879	0.1264	0.5723	0.1479
1300-1500	0.8452	0.0361	0.7796	0.0431
1500-1700	0.9267	0.0248	0.9041	0.0288
1700-1900	0.8347	0.0373	0.7584	0.0451
1900-2100	0.8363	0.0371	0.7940	0.0416
2100-2300	0.7563	0.1116	0.6894	0.1260
2300-2500	0.8697	0.0816	0.8232	0.0950

Table 6. 47 synergic interval Partial Least Square regression model evaluation of carbohydrate with some selected wavelengths

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
1500-1700, 1900-2100, 2300-2500	0.9659	0.0079	0.9536	0.0089
1500-1700, 2300-2500	0.9345	0.0338	0.9145	0.0389

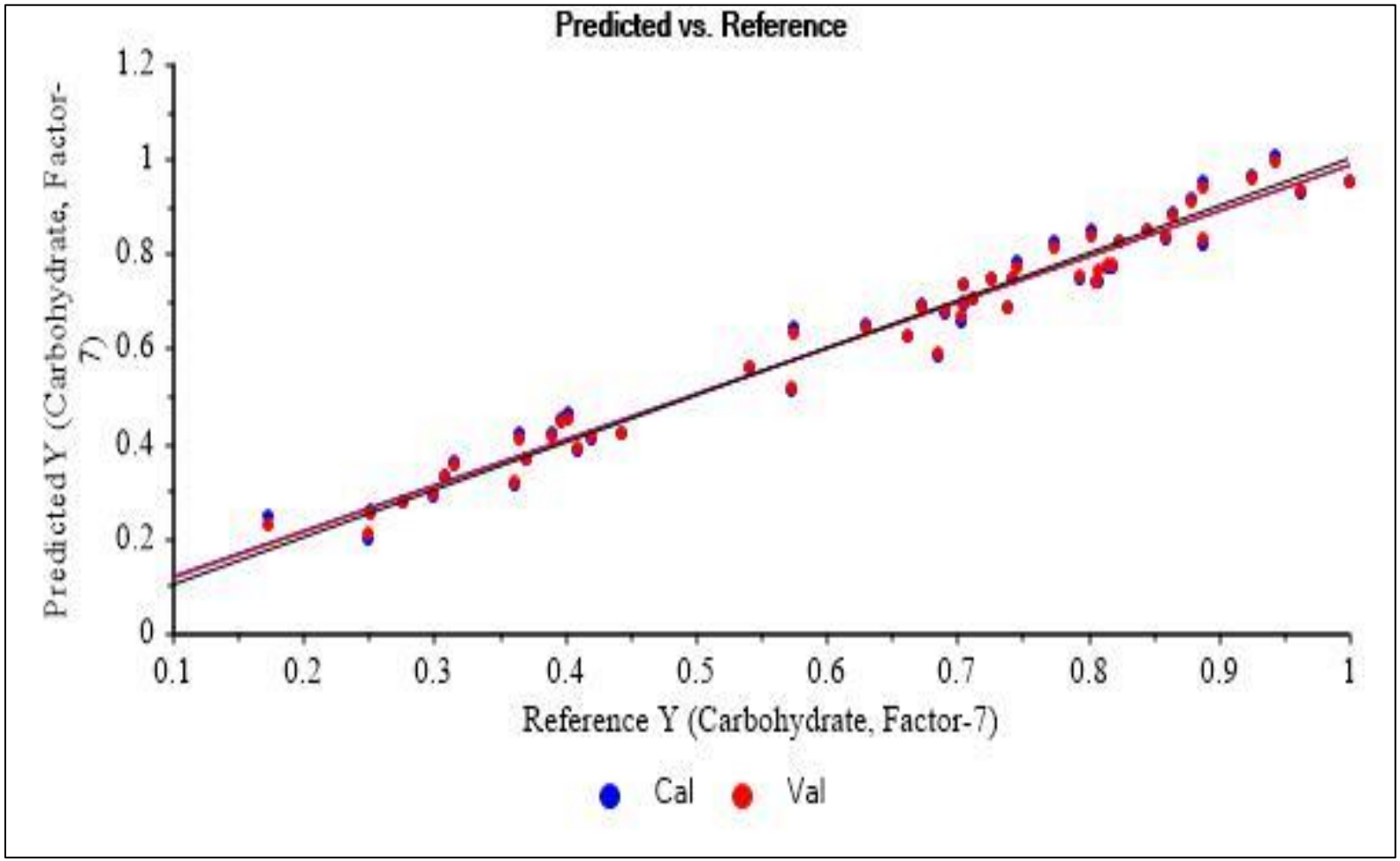


Figure 6. 21 Scatter plot of mango varieties for carbohydrate content synergic interval Partial Least Square regression in selected wavelength range for calibration and validation

Table 6. 48 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for carbohydrate with Principal Component Regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.2025	0.2019	0.1009	0.2144
	3	0.2029	0.2001	0.0008	0.2260
	4	0.3909	0.1765	0.1833	0.2044
	5	0.4278	0.1710	0.1749	0.2054
	6	0.4401	0.1692	0.1270	0.2113
	7	0.4402	0.1660	0.0674	0.2184
MSC	2	0.0013	3.6457	NA	3.9167
	3	0.0388	3.5767	NA	5.0936
	4	0.0775	3.5039	NA	4.9012
	5	0.2053	3.2520	NA	4.2216
	6	0.2554	3.1479	NA	3.8377
	7	0.2865	3.0814	0.0323	3.5888
2^{nd} order derivative	2	0.0186	3.6748	NA	3.9092
	3	0.0216	3.6691	NA	4.0063
	4	0.0647	3.5875	NA	3.9999
	5	0.0724	3.5726	NA	4.1693
	6	0.1198	3.4802	NA	4.5633
	7	0.2836	3.1395	NA	4.7968
MSC + 2^{nd} order derivative	2	0.1499	3.4200	0.0407	3.6331
	3	0.1533	3.4132	0.0631	3.7037
	4	0.1542	3.4115	NA	3.8270
	5	0.2170	3.2822	NA	3.7533
	6	0.3038	3.0905	0.0523	3.6110
	7	0.3274	3.0421	0.0283	3.6566

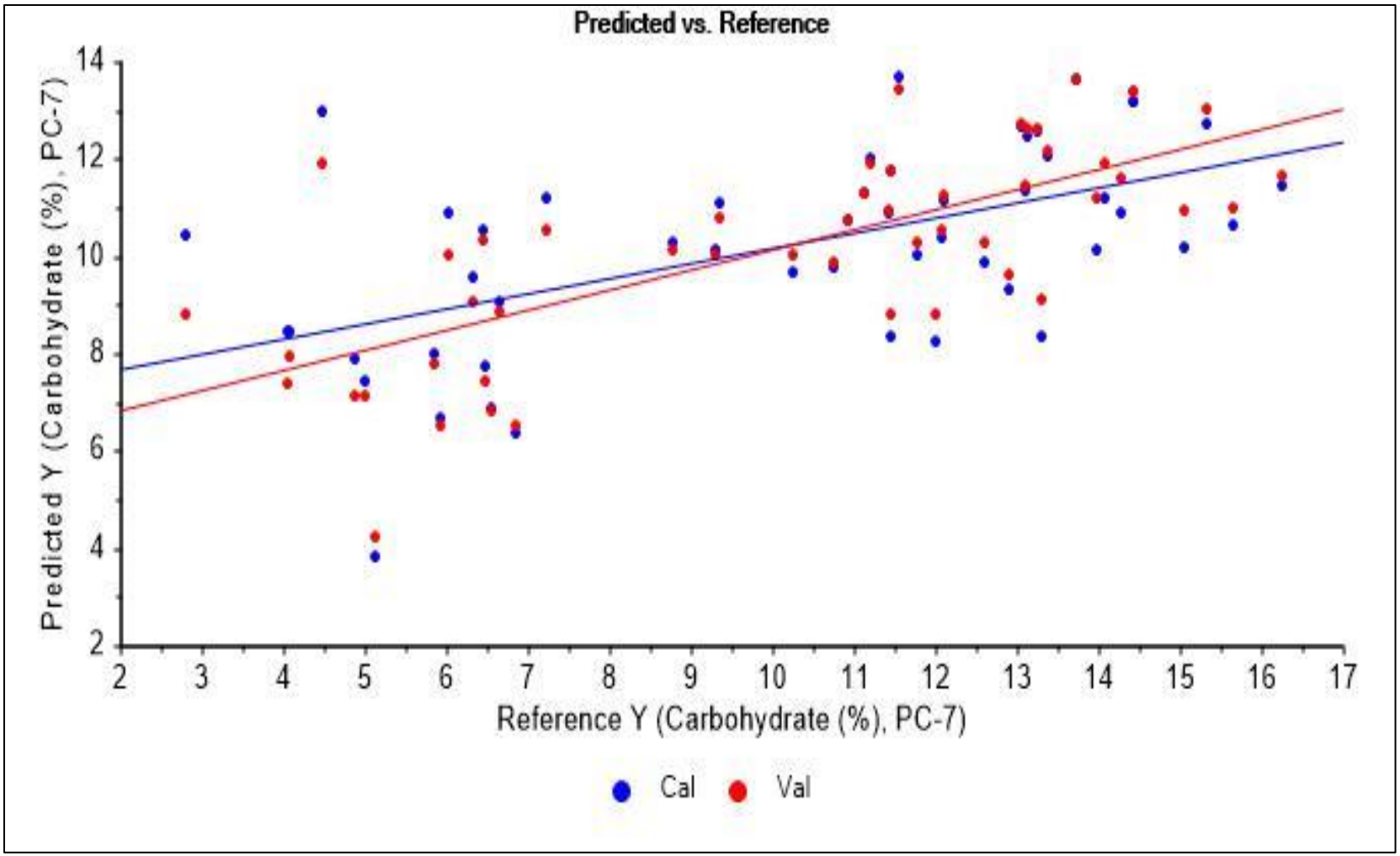


Figure 6. 22 Scatter plot of mango varieties for carbohydrate content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation Principal Component Regression

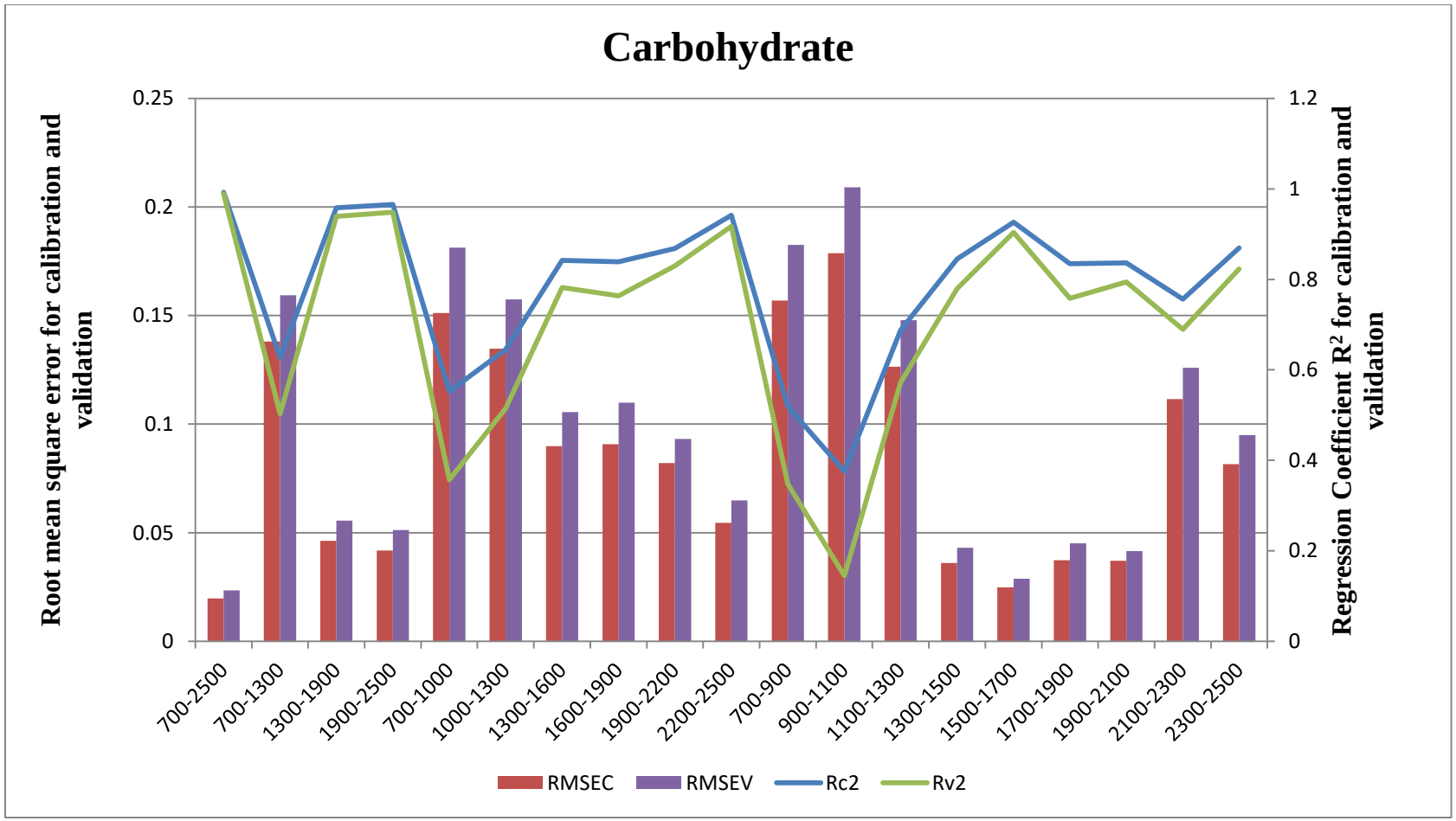


Figure 6. 23 Variation of regression coefficient and root mean square error for calibration and validation with different wavelength range

Table 6. 49 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for total soluble solid with Partial Least Square regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.1356	0.1224	0.0388	0.1291
	3	0.1412	0.1220	NA	0.1319
	4	0.1729	0.1197	NA	0.1325
	5	0.2208	0.1162	NA	0.1317
	6	0.2854	0.1113	0.0240	0.1300
	7	0.3204	0.1085	0.0557	0.1279
MSC	2	0.0540	0.1280	NA	0.1361
	3	0.1492	0.1214	NA	0.1320
	4	0.2071	0.1172	0.0452	0.1286
	5	0.2515	0.1139	0.0622	0.1275
	6	0.3272	0.1080	0.1379	0.1222
	7	0.3943	0.1024	0.1852	0.1188
2^{nd} order derivative	2	0.6640	0.0763	0.6301	0.0800
	3	0.7640	0.0639	0.7282	0.0686
	4	0.8266	0.0548	0.7893	0.0604
	5	0.9009	0.0414	0.8752	0.0465
	6	0.9365	0.0331	0.9176	0.0377
	7	0.9569	0.0273	0.9411	0.0319
MSC + 2^{nd} order derivative	2	0.6611	0.0766	0.6267	0.0804
	3	0.7665	0.0636	0.7312	0.0682
	4	0.8275	0.0546	0.7910	0.0601
	5	0.9035	0.0408	0.8774	0.0461
	6	0.9372	0.0329	0.9188	0.0375
	7	0.9576	0.0270	0.9426	0.0315

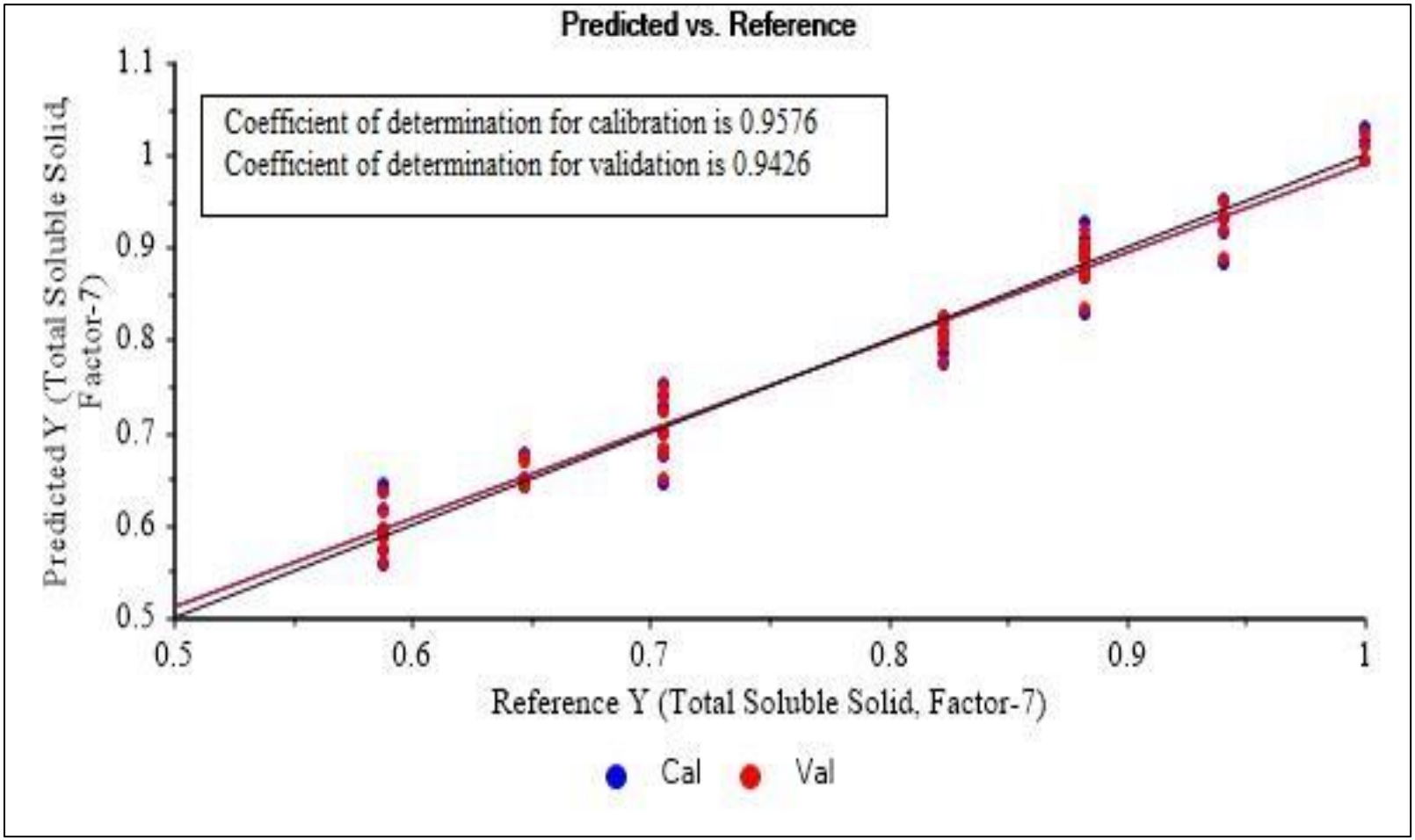


Figure 6. 24 Scatter plot of mango varieties for total soluble solid after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation Partial Least Square regression

Table 6. 50 interval Partial Least Square regression model evaluation of total soluble solid within the range of 700-1300 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.1025	0.1247	NA	0.1349
3	0.1590	0.1207	0.0006	0.1316
4	0.1918	0.1183	0.0028	0.1314
5	0.3959	0.1023	0.2450	0.1142
6	0.6046	0.0828	0.4849	0.0945
7	0.7436	0.0666	0.6557	0.0772

Table 6. 51 interval Partial Least Square regression model evaluation of total soluble solid within the range of 1300-1900 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.2886	0.1110	0.1928	0.1183
3	0.6436	0.0786	0.5858	0.0847
4	0.7741	0.0625	0.7262	0.0689
5	0.8687	0.0477	0.8342	0.0536
6	0.9209	0.0370	0.8894	0.0437
7	0.9554	0.0277	0.9374	0.0329

Table 6. 52 interval Partial Least Square regression model evaluation of total soluble solid within the range of 1900-2500 nm

Factors	R_c^2	RMSEC	R_v^2	RMSEV
2	0.5361	0.0896	0.4809	0.0948
3	0.7057	0.0714	0.6700	0.0756
4	0.7933	0.0598	0.7584	0.0647
5	0.8803	0.0.455	0.8530	0.0504
6	0.9043	0.0407	0.8736	0.0468
7	0.9343	0.0337	0.9148	0.0384

Table 6. 53 interval Partial Least Square regression model evaluation of total soluble solid with the interval of 300 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-1000	0.4451	0.0980	0.2642	0.1129
1000-1300	0.6523	0.0776	0.5393	0.0893
1300-1600	0.9102	0.0394	0.8729	0.0469
1600-1900	0.9020	0.0412	0.8664	0.0481
1900-2200	0.9180	0.0376	0.8844	0.0447
2200-2500	0.8680	0.0478	0.8262	0.0548

Table 6. 54 interval Partial Least Square regression model evaluation of total soluble solid with the interval of 200 nm

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
700-900	0.3895	0.1028	0.2119	0.1169
900-1100	0.3072	0.1096	0.0748	0.1266
1100-1300	0.6934	0.0729	0.5630	0.0870
1300-1500	0.8056	0.0580	0.7288	0.0685
1500-1700	0.7750	0.0624	0.6910	0.0731
1700-1900	0.8695	0.0475	0.8257	0.0549
1900-2100	0.8455	0.0517	0.8067	0.0578
2100-2300	0.8369	0.0531	0.7848	0.0610
2300-2500	0.8237	0.0552	0.7673	0.0635

Table 6. 55 synergic interval Partial Least Square regression model evaluation of total soluble solid with some selected wavelengths

Wavelength	R_c^2	RMSEC	R_v^2	RMSEV
1300-1500, 1700-1900	0.9513	0.0245	0.9344	0.0249
1700-1900, 2300-2500	0.9242	0.0300	0.9012	0.0575

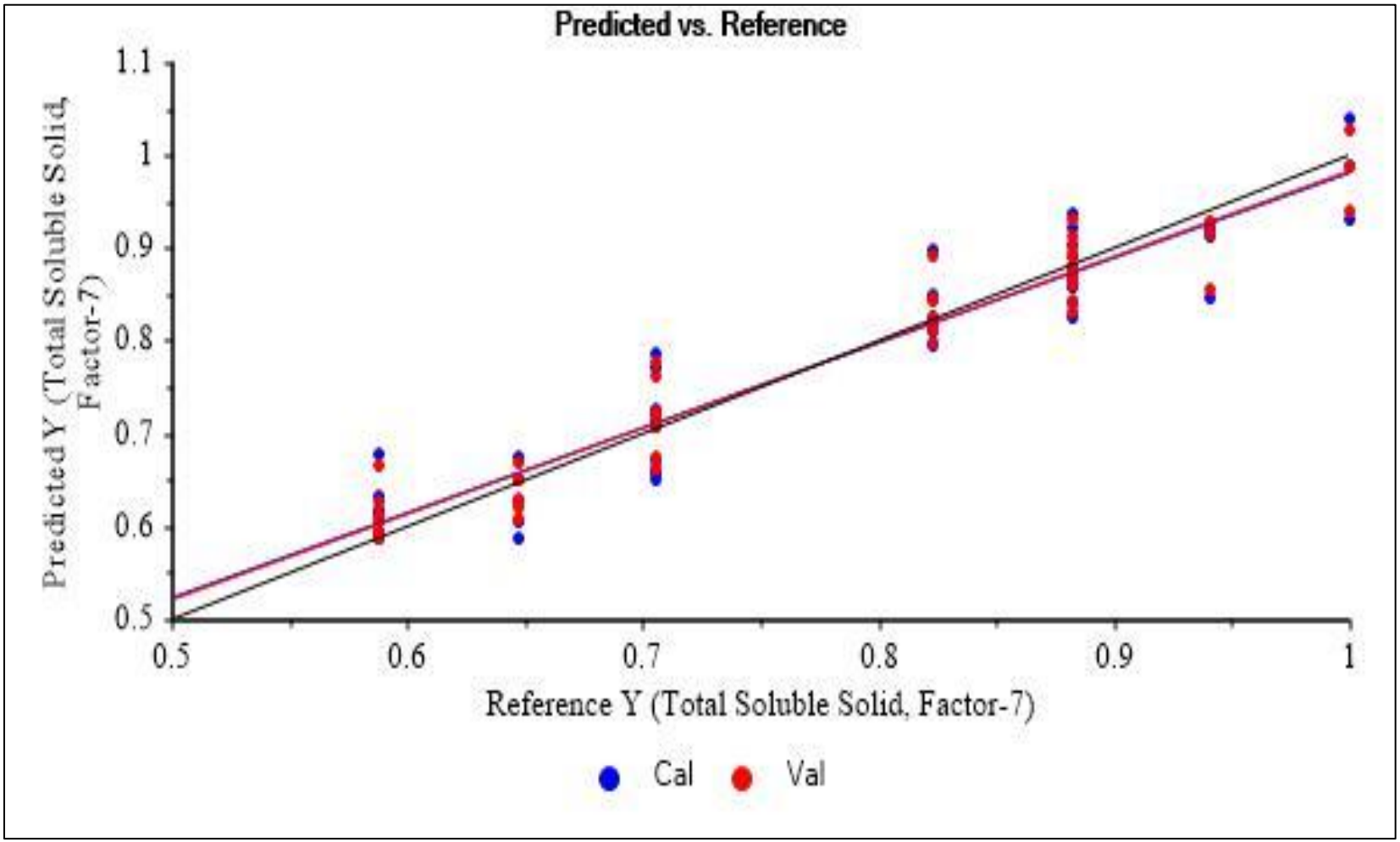


Figure 6. 25 Scatter plot of mango varieties for total soluble solid synergic interval Partial Least Square regression in selected wavelength range for calibration and validation

Table 6. 56 Effect of pre-processing method on the spectral data of total range of 700-2500 nm for total soluble solid with Principal Component Regression

Preprocessing techniques	Factors	R_c^2	RMSEC	R_v^2	RMSEV
Original spectra	2	0.0400	0.1290	NA	0.1375
	3	0.1343	0.1225	NA	0.1325
	4	0.1359	0.1224	NA	0.1357
	5	0.1429	0.1219	NA	0.1378
	6	0.1524	0.1212	NA	0.1402
	7	0.1527	0.1210	NA	0.1450
MSC	2	0.0910	2.1662	NA	2.2719
	3	0.1018	2.1532	NA	2.2942
	4	0.1032	2.1516	NA	2.4097
	5	0.1449	2.1010	NA	2.4013
	6	0.1795	2.0580	NA	2.6094
	7	0.1942	2.0395	NA	2.5176
2^{nd} order derivative	2	0.0625	2.1797	NA	2.2920
	3	0.0785	2.1641	NA	2.3321
	4	0.0825	2.1595	NA	2.3515
	5	0.1383	2.0927	NA	2.3362
	6	0.1387	2.0922	NA	2.3943
	7	0.1397	2.0911	NA	2.4803
MSC + 2^{nd} order derivative	2	0.0155	2.2369	NA	2.4051
	3	0.0208	2.2308	NA	2.4472
	4	0.0352	0.2144	NA	2.4652
	5	0.1251	2.1087	NA	2.819
	6	0.1287	2.1043	NA	2.4398
	7	0.1297	2.1031	NA	2.5137

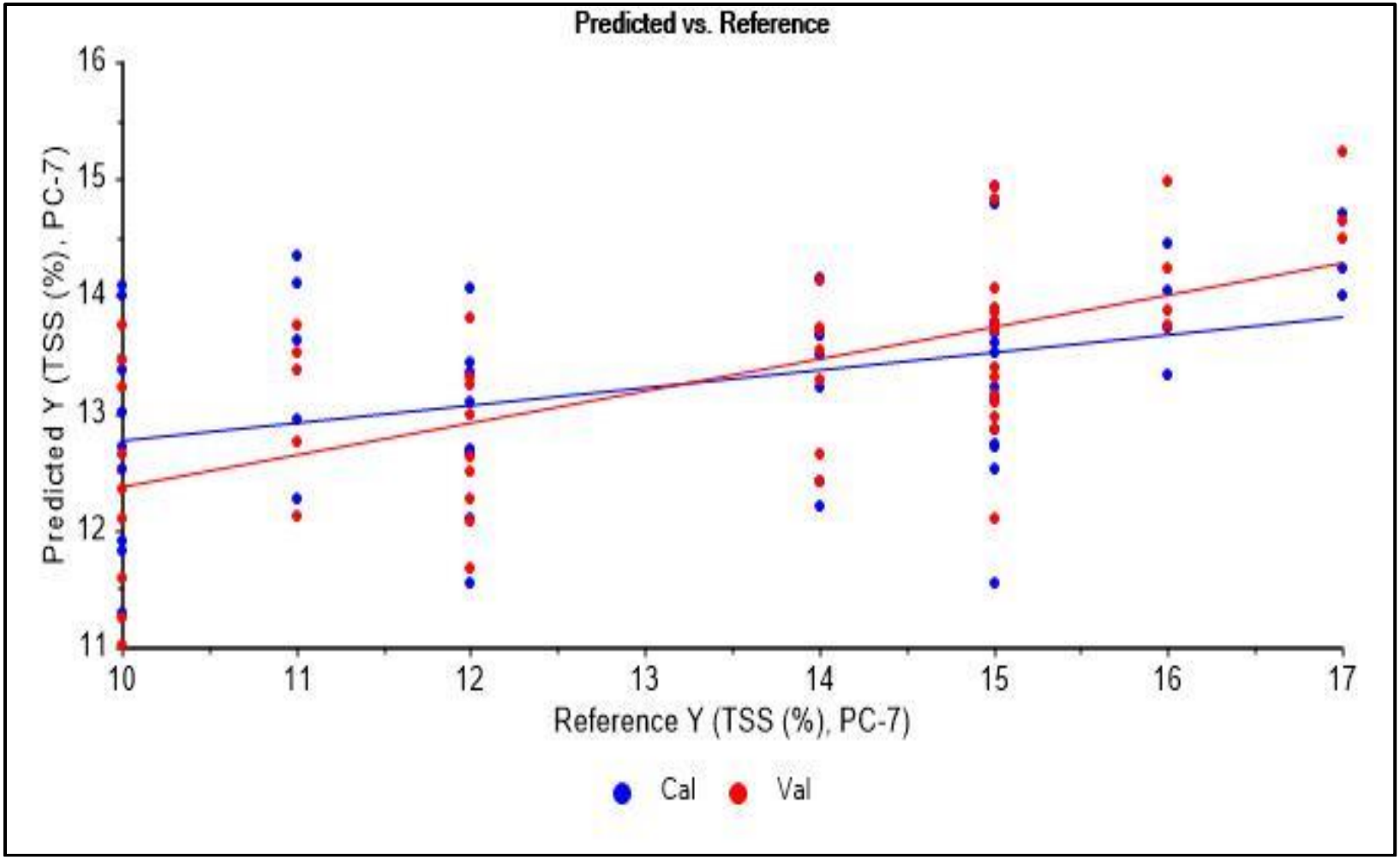


Figure 6. 26 Scatter plot of mango varieties for total soluble solid content after Multiplicative Scatter Correction and second order derivative in wavelength range of 700-2500 nm for calibration and validation Principal Component Regression

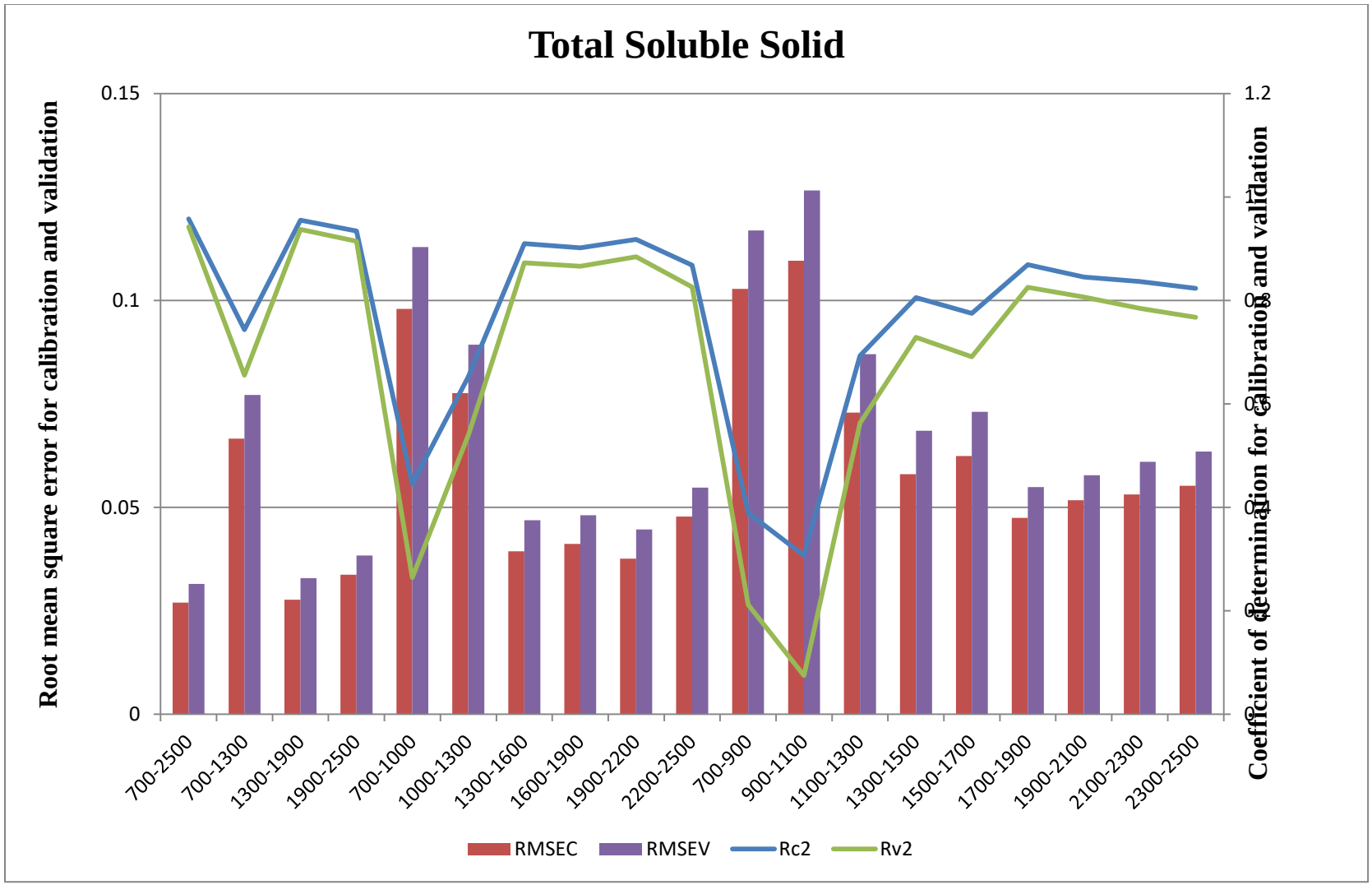


Figure 6. 27 Variation of regression coefficient and root mean square error for calibration and validation with different wavelength range

6.3.4. Prediction of the Quality Attributes of Mango

Prediction of all the quality parameters of mango – moisture, crude fat, crude protein, crude fiber, pH, carbohydrate and TSS has done and the deviation between predicted value and reference value has determined. For all the case it has observed that the reference value of every parameter and the predicted value of every parameter are almost same and the minimum deviation has shown (Table 6.57). So, this can lead a promising conclusion for the application of NIR technology to predict the quality parameter of mango pulp in a non-destructive manner.

Table 6. 57 Table for the predicted values of the quality parameter of mango

Parameters	Predicted value	Reference value	Deviation
Moisture	0.9098	0.9076	0.0077
Crude fat	1.2526	0.9835	0.0537
Crude protein	0.9288	0.9724	0.0535
Crude fiber	1.0733	0.7622	0.0336
pH	0.8935	0.9374	0.0227
Carbohydrate	0.7573	0.6671	0.0796
Total soluble solid	0.6366	0.8529	0.0254

6.4. Conclusion

This study has concluded with a promising result of the application of NIR technology. NIR spectral data along with analytical data has been used for the regression model development. For modeling, raw spectral data has not given the satisfactory results, whereas the application of pre-processing (MSC, second order derivative and the combination of both) has modified the spectra. Two different chemometric approach (PLS and PCR) has been used and a comparison of the results has been done. It has concluded that PLS is giving the best results in terms of errors and coefficient values for the fresh mango samples. Furthermore iPLS and siPLS have used to reduce the spectral data points. As results, regression models for quantitative and qualitative analysis has developed to predict the quality parameters of mango like moisture, crude fat, crude fiber,

crude protein, pH, carbohydrate and TSS. Prediction results have developed to check the accessibility of the validated regression model.

`References

1. J. Chave, C. Andalo, S. Brown, M. A. Cairns, J. Q. Chambers and D. Eamus, H. Fölster, F. Fromard, N. Higuchi, T. Kira and J. P. Lescure, B. W. Nelson, H. Ogawa, H. Puig, B. Riera and T. Yamakura, *Oecologia*, **145**, 87-99 (2005).
2. Y. Ye, H. Guo and X. Sun, *Biosens. Bioelectron.*, **126**, 389-404 (2019).
3. M. E. Raypah, A. F. Omar, J. Muncan, M. Zulkurnain and A. R. Abdul Najib, *Molecules*, **27**, 2324 (2022).
4. G. Dryden, *Near infrared spectroscopy: Applications in deer nutrition* (*RIRDC Pub*, 2003) (W03/007).
5. H. Nguyen, F. Lagarde, G. Louarn and P. Daniel, *Talanta*, **167**, 436-441 (2017).
6. D. A. Burns and E. W. Ciurczak, *Handbook of near-infrared analysis* (CRC press, 2007).
7. E. W. Ciurczak, B. Igne, J. J. Workman and D. A. Burns, *Handbook of near-infrared analysis* (CRC press, 2021).

Chapter-7: Conclusion

Chapter-7

Conclusion

The research work focused on determination of physicochemical parameters of Indian mango pulp namely – peel content, pulp content, kernel weight, all the diameters of intact mango (major, minor, intermediate), sphericity, surface area, size of intact mango, carbohydrate, crude protein, crude fiber, crude fat, moisture, ash content, pH, total soluble solid (TSS) and energy value. A variation in all these parameter values has been noticed and the reason of this variation might be introduced due to the different nature of soil, different harvest methods and variation of temperature. Secondly, after Kearl Pearson correlation has performed, a correlation between all the physical and chemical parameters was observed. Some correlations were strong enough and some were poor. The nature of correlation was analyzed with the correlation coefficients (r) and it was observed that energy, moisture and carbohydrate content showed the finest correlation among the all chemical attributes present in mango pulp. Also, it was observed that pulp content, weight of intact mango fruit, size and surface area of the mango fruit showed the best correlation among all physical attributes of mango. A regression equation has been developed to justify the linear relations among all the correlated parameters. The result of viscosity has shown a wide range of values, which depicted that there was variation in the viscosity value of Indian mangoes. Some varities are having very high viscous pulp, whereas, some have very low viscous in nature. The variation in viscosity might be possible for the variation of temperature.

Near infrared and Fourier Transform infrared spectroscopy has been applied to check the presence of organic molecules in the mango. This is a primary data to estimate the quality parameters (water content, carbohydrate, TSS, pH, crude fat, crude protein and crude fiber) of mango pulp along with the chemometrics. Principal Component Analysis (PCA) was employed using singular value decomposition algorithm to classify the mango

samples on basis of origin. Samples were taken from five different states of India these are Bihar, Maharashtra, Punjab, Uttar Pradesh and West Bengal. The results were indicated the classification of mango on basis of origin is hardly possible for the total datasets. But after reduction of some data sets, the samples of Uttar Pradesh and West Bengal were classified in a better way. But, among all the parameters that physical attribute, chemical attribute and spectral (NIR and FTIR) attributes, the samples were well classified with the FTIR data of the samples of west Bengal and Uttar Pradesh. So, these findings can conclude that the varities of mangoes are hardly different in nature with each other. Although there is a variation in physical and chemical parameters of different varities, but the range is not too large to classify the samples in terms of varities of mango.

Next aspect was to predict the qualitative parameters of mango using near infrared spectroscopy (700-2500 nm) together with chemmometric approach - PLS regression. Qualitative and quantitative analysis was performed with the chemmometric study to estimate the quality parameters namely carbohydrate, crude protein, crude fiber, crude fat, moisture, pH and TSS. It was observed that model showed good accuracy with R_c^2 and R_v^2; results more than 0.9 for all the parameters. The purpose of final model was to predict the parameters with minimum wavelengths, so that the prediction model will be cost and time effective. For this purpose wavelength selection was performed with interval PLS and synergic PLS method. These methods are manually reduced the wavelengths and obtained the best results of R^2. In the present research for moisture, crude protein, crude fiber, carbohydrate, pH and TSS total 3599 spectral data point was reduced into 800 data points and for crude fat the total data points was reduced into 2000 data points. The accuracy of the model has been justified with the higher values of coefficient of determination and the lower values of the root mean square error values for calibration and validation, which is near to zero. The results revealed that NIR spectroscopy along with chemometrics was an efficient method to predict the quality of mangoes with green and non-hazardous way. NIR calibration method was reduced the process time and the costing of the analysis. Quantity of sample also was needed very

small in this case. Prediction models were used determine almost all the physicochemical properties, with maximum possible number of data point reduction, only in case of crude fiber the data point reduction was less. Despite of this the results of the present research are promising enough for the quantitative and qualitative analysis of mangoes collected from five different states (Bihar, Maharashtra, Punjab, Uttar Pradesh and West Bengal) of India.

Chapter-8: Future Scope

Chapter-8

Future scope

- Portable NIR instrument can be designed using selected wavelengths for qualitative and quantitative analysis of mango to reduce the cost, estimation time and to enhance the accuracy of the results.

- Aquaphotomics can be applied on the same data to estimate the water content present in the mango fruit.

- There are other quality parameters of mango like vitamin A content, vitamin C content, vitamin B content, titratable acidity of mango which can be determined using the standard AOAC method and also regression model can be developed with selected wavelengths for estimation of these parameters with green and non-destructive way.

- There are many more local varities of mango for different states of India are available in different mango farms and local markets, but they are having very less commercial importance. To enhance their commercial scale a large project can be designed to estimate the quality parameters of mango for each varities. More than, fifty samples can be taken from every state and chemmometric can be applied on the large scale data.

- There are some other spectroscopic techniques like Fourier Transform infrared spectroscopy, UV-Vis spectroscopy which can be explored for the same study.

List of Publication

Published Articles

1. **Majumdar, A. D.**, Kaur, A., Munjal, N., Kamboj, U. and Pramanik, T. (2020), "Detection of different adulteration from milk with non-destructive approach (Near infrared spectroscopy): A short review", *American Journal of Physical Sciences and Applications*, 1(3), 14-15.
2. **Majumdar, A. D.**, Kamboj, U. and Munjal, N. (2022), "Physicochemical Properties of Mango Pulp and Mid-infrared Characterization", *Journal of Physics: Conference Series (*IOP Publishing), 2267(1), 012035.
3. **Majumdar, A. D.**, Munjal, N. and Kamboj, U. (2022), "Characterization of *Pennisetum Glaucum* (Pearl Millet)", *Journal of Physics: Conference Series* (IOP Publishing), 2267(1), 012020.
4. **Majumdar, A. D.**, Kamboj U. and Munjal N. (2023), "Geographical Classification with Principle Component Analysis Related to Chemical attributes and Non-destructive Assessment of Sweetness of Indian Mangoes using chemometrics" *Annals of Agri Bio Research* , 28(1), 129-1

Book Published

1. **Agnibha Das Majumdar**, Uma Kamboj and Neha Munjal, "*Qualitative and quantitative characterization of mango*" Notion press; 1st edition (22 July, 2022), pp-1-88, ISBN-979-8887722986.

Accepted Articles

1. **Agnibha Das Majumdar**, Uma Kamboj, Neha Munjal. "Chemical composition of mango pulp and the comparison of two vibrational spectra" *AIP Conference proceedings* (SJR-0.189) in ICMET- 2022 held on 18-19 February, 2022 (Accepted).

Communicated Articles

1. **Majumdar, A. D**., Kamboj, U and Munjal, N., (2022), "Characterization and geographical classification of physical and spectral attributes of mango (*Mangifera indica*) with principal component analysis", *South African Journal of Botany* (Communicated).
2. **Majumdar, A. D**., Kamboj, U and Munjal, N., (2023), "Green Method for Chemical Assessment of Indian Mangoes Using Chemometrics", *Rasayan Journal of Chemistry* (Communicated).
3. **Majumdar, A. D**., Kamboj, U and Munjal, N., (2023), "Fourier Transform IR Spectroscopy: A Qualitative Tool to Analyze Fruit Peel", *Journal of Tropical Agriculture* (Communicated)

List of conference/ workshop/ short-term course attended

1. Presented oral presentation entitled as, “Near infrared spectroscopy an important tool to detect the chemical properties of Mango (*Mangifera indica*): a mini review” in International Web-Conference on Advanced Nanostructured Materials for Energy Generation, Storage, and Smart Applications (SPMCANM-2020) on 9th and 10th October, 2020 organized by department of Physics, Shardabai Pawar Mahila Arts, Commerce and Science College, Shardanagar, Baramati, Pune.
2. Presented oral presentation entitled as, “The impact of non-destructive approaches with the Chemometric analysis on the external and internal quality attributes of Mango (*Mangifera indica*): a review” in International Multidisciplinary Conference in Current Research Trends (IMCCRT-2020) on 19th and 20th September, 2020 organized by Research Circle Team.
3. Presented oral presentation entitled as, “Pearl Millet: A Review of its Physical and Chemical Properties” in International Multidisciplinary Conference in Current Research Trends (IMCCRT-2020) on 19th and 20th September, 2020 organized by Research Circle Team.
4. Presented oral presentation entitled as, “Physicochemical Properties of Mango Pulp and Mid-infrared Characterization” in Recent Advances in Fundamental and Applied Sciences (RAFAS-2021) on 25th and 26th June, 2021 organized by School of Chemical Engineering and Physical Sciences, Lovely Professional University, Phagwara, Punjab.
5. Presented oral presentation entitled as, “Chemical Composition of Mango pulp and the comparison of two vibrational spectra” in International Conference on Materials of Emerging Technologies (ICMET-2022) on 18th and 19th February, 2022 organized by Department of research impact and outcome, Division of Research and Development, Lovely Professional University, Phagwara, Punjab.

6. Presented oral presentation entitled as, “Characterization of Mango With Physicochemical and spectral attributes” in International Conference on Integrated Approaches in Science and Technology for Sustainable Future” (IASTSF-2022) on 28th February to 1st March, 2022 organized by J C Bose University of Science and Technology, YMCA, Faridabad, Haryana.
7. Presented oral presentation entitled as, “Characterization and geographical classification of physical and spectral attributes of mango (*Mangifera indica*) with Principal Component Analysis” in International Conference on Advances in Agriculture Technology and Allied Sciences (ICAATAS-2022) on 4th and 5th June, 2022 organized by, MS Swaminathan School of Agriculture, Centurion University of Technology and Management, Parlekhamundi, Odisha and Association of Rice Research Workers (ARRW), NRRI.
8. Presented oral presentation entitled as, “Green method for chemical assessment of Indian mangoes using chemometrics” in Recent Advances in Fundamental and Applied Sciences (RAFAS-2023) on 24th and 25th March, 2023 organized by School of Chemical Engineering and Physical Sciences, Lovely Professional University, Phagwara, Punjab.

Achievements during Ph.D.

1. Presented research papers at more than 25 international/national conferences (oral/poster) and visited Bangladesh with a travel grant for 3 days (Dec 2019) for oral presentation at Chattagram University of Engineering and Technology, Bangladesh (ICPSDT-2019).
2. Received 3 rd Prize for Oral presentation in the "International Conference on Advances in Agriculture Technology and Applied Sciences" (ICAATAS-2022), organized by Centurion University, Odisha.
3. Delivered lecture on "Spectroscopy" on the platform of ISTEM (Indian Science, Technology and Engineering facilities Map) - India, an online portal launched by the Honorable Prime Minister of India Shri Narendra Modi.