UNIVERSITY OF CAPE COAST

DEVELOPING A NOVEL ONSITE DETECTION TECHNOLOGY BY USING CHEMOMETRICAL ANALYSIS OF HAND-HELD NEAR-INFRARED SENSOR TECHNIQUE FOR ASSESSING COFFEE QUALITY

VIDA GYIMAH BOADU

2024

Vida Gyimah Boadu
University of Cape Coast

UNIVERSITY OF CAPE COAST

DEVELOPING A NOVEL ONSITE DETECTION TECHNOLOGY BY USING CHEMOMETRICAL ANALYSIS OF HAND-HELD NEAR-INFRARED SENSOR TECHNIQUE FOR ASSESSING COFFEE QUALITY

BY

VIDA GYIMAH BOADU

(AG/DHP/19/0002)

Thesis submitted to the Department of Agricultural Engineering of the School of Agriculture, College of Agriculture and Natural Sciences, University of Cape Coast, in partial fulfilment of the requirements for the award of Doctor of Philosophy degree in Food and Postharvest Technology

JUNE, 2024

DECLARATION

Candidate's Declaration

I hereby declare that this thesis is the result of my own original research and that no part of it has been presented for another degree in this university or elsewhere.

Candidate's Signature ……Date ……

Name: ……

Supervisors' Declaration

We hereby declare that the preparation and presentation of the thesis were supervised in accordance with the guidelines on supervision of thesis laid down by the University of Cape Coast.

Principal Supervisor's Signature …… Date ……

Name: ……

Co-Supervisor's Signature …… Date …….

Name: ……

ABSTRACT

This study aims to apply handheld infrared spectral technique to develop a predictive model for the identification of adulterants and rapid estimation of the quality of coffee products. For the classification of coffee varieties, the novel potable NIR spectrometer combined with multivariant qualitative algorithms gave 98.8%, 99.72% and 99.22% identification for raw, roasted and roasted powdered coffee, respectively. Also, in the classification of Africa coffee types, it gave 99.76%, 99.78% and 99.88% identification for raw, roasted and roasted powdered coffee, respectively. Qualitatively, FD-LDA performed better with 97.78% and 100% in both calibration and prediction sets in the determination of coffee husk in roasted coffee powder. Quantitatively, in the detection of coffee adulteration, SPA-PLS model had the best results with R=0.9711 and 0.9897 in both calibration and prediction sets respectively. The novel handheld spectroscopy could be employed for the discrimination of coffee varieties and African robusta coffee in three forms (raw, roasted and powder) and quantification of coffee husk in coffee. With 10% occurrence frequencies, two fungi, *Aspergillus niger* and *flavus* were found in commercially sold powder coffee in some of the major markets in Ghana with acrylamide levels below the benchmark threshold (400ug/kg) set by the European Commission. The proximate analysis conducted on the commercially sold coffee powder revealed high moisture and ash attributed to a substantial amount of impurities in the coffee samples. Furthermore, minerals namely nitrogen, phosphorus, potassium and magnesium were found in the coffee powders.

LIST OF PUBLICATIONS BY CANDIDATE

1. Boadu, V. G., Teye, E., Amuah, C. L., & Sam-Amoah, L. K. (2022). Rapid authentication of coffee bean varieties of different forms by using a pocket-sized spectrometer and multivariate data modelling. *Analytical Methods*, *14*(46), 4756-4766.

2. Boadu, V. G., Teye, E., Amuah, C. L., Lamptey, F. P., & Sam-Amoah, L. K. (2023). Portable NIR Spectroscopic Application for Coffee Integrity and Detection of Adulteration with Coffee Husk. *Processes*, *11*(4), 1140.

3. Boadu, V. G., Teye, E., Amuah, C. L., & Sam-Amoah, L. K. (2024). Rapid classification of African geographical coffee types by handheld NIR spectroscopic method. (Accepted for publication in Foods Journal ISSN2304-8158)

ACKNOWLEDGMENT

The grace and health that God has given me each day to complete this study is a blessing for which I am incredibly grateful. I would like to express my sincere gratitude to Prof. Ernest Teye, who served as my main supervisor and advisor, for his positive feedback, outstanding direction, suggestions, and constant encouragement throughout my time at the University. He is very knowledgeable, organized, and competent, all of which have been helpful to me. I also want to thank my co-supervisor, Prof. Livingstone Sam-Amoah, for his support and encouragement.

I also would like to express my gratitude for the support and inspiration I received from Prof. Ernest Ekow Abano and Prof. R. S. Amoah, as well as the entire staff of the Department of Agricultural Engineering. I appreciate the help of Dr. Charles L.Y. Amuah, Francis Padi Lamptey, Edward Ken Essuman, Emmanuel Korsah, Regina Nyokeh, and Welbeck for their valuable assistance throughout the research work.

I would like to express my sincere gratitude to my husband, Mr. Maxwell Osei Boadu, for his constant encouragement, perseverance, prayers, and encouragement.

DEDICATION

I dedicate this work to my beloved Husband, Mr. Maxwell Osei Boadu and lovely children; Yaw, Kwasi, and Akua

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

1st- Der – First Derivative

2nd-Der – Second Derivative

3rd Der – Third Derivative

ANN – Artificial Neural Network

AS- Autoscaling

biPLS- Backward Interval Partial Least Square

BP-AdaBoost- Back Propagation Adaptive Boosting

BP-ANN- Back propagation artificial neural network

DC- Detrend correction

DOSC- Direct Orthogonal Signal Correction

DPLS- Discriminate Partial Least Square

EMSC- Extended Multiplicative scatter correction

Ga- PLS- Genetic Algorithm Interval Partial Least Square

HCA- Hierarchical Cluster Analysis

iPLS- Interval Partial Least Square

IPW-PLS- Iterative predictor weighing Partial Least Square

ISE-PLS- Interactive stepwise elimination Partial Least Square

KNN- K-nearest neighbour

LDA- Linear Discrimination Analysis

MC- Mean centering

MLR- Multiple Linear Regression

MP-PLS – Multiblock Partial Least Square

MSC- Multiplicative scatter correction

NN- Neural Network

OCKNN- One-class K-nearest neighbour

OCSVM- One-class Support vector machine

OLS- Ordinary Least- Squares

PCA- Principal Component Analysis

PLS- Partial Least Square

PLSDA- Partial Least Square Discrimination Analysis

PLSR- Partial Least Square Regression

R- Correlation coefficient

RF- Random Forest

RMSECV- Root mean square error cross validation

RMSEP- Root mean square error prediction

RPD- Ratio performance deviation

SELECT- Stepwise Orthogonalization of predictors

SG- Savitsky-Golay

SIMCA- Soft Independence Modeling Class Analogy

SiPLS – Synergy Interval Partial Least Square

SPA-PLS – Successive Projections Algorithms Interval Partial Least Square

SVM- Support vector machine

UVE-PLS- Uninformation variable elimination

CHAPTER ONE

INTRODUCTION

1.1 Background to the Study

Coffee fruit is a berry or cherry stone fruit that consists of a tough smooth pericarp, a soft yellowish fibrous mesocarp, and a yellowish endocarp that covers the endosperm (coffee bean). The pericarp is green in the unripe stage but turns to deep red or red-violet at the ripped stage in maturation (Esquivel & Jimenez, 2012). It takes about five years to harvest raw coffee cherries, which are made up of two beans encased in a hull and surrounded by pulp (Arya & Rao, 2007). Coffee is defined as a dried seed of a coffee plant whether such seed is in its raw state or roasted and ground state in UK Coffee and Coffee Products Regulations (SI 1420). The key by-products of coffee are the pulp, peel, and husk which is about 45% of the cherry which are also valuable since caffeine and polyphenols are extracted from them (Esquivel & Jimenez, 2012). The coffee tree belongs to the family Rubiaceae and genera coffea. It was discovered in Ethiopia and sent to Arab and Western Europe in the sixteenth and seventeenth centuries respectively. After three to four years, the coffee tree begins to bloom, and six years afterward, it is completely productive. After 10 to 15 years, the maximum yield is reached, and the plant continues to produce fruit for about 40 years. During cultivation, the tree is pruned to a height of 2 – 2.5 meters to enable harvest. The tree is an evergreen with fibrous short–stem leaves and snowy flowers with a Jasmin-like fragrance (Bondesson, 2015). Coffee plantations are located in geographical bands encompassing the world called “coffee belts” found between latitude 20ºN and

20ºS. The band encircles the center and Southern part of America, Asia, and Africa and is confined by the tropics of Capricorn and Cancer which provide an ideal climatic condition for growing coffee plants (Giraudo et al., 2019). Robusta and Arabica coffee are two species that account for 99% of the world's bean production, making them both economically significant on a global scale. Arabica coffee grows well in high altitudes from 1300 and 2000m. Robusta (*Coffea canephora*) is extensively dispersed in tropical Africa and adapts well in altitudes below 1000m (Chu, 2012). Arabica coffee is more highly priced and appreciated by consumers for its flavor than Robusta but the latter is resistant to pests and diseases.

In Africa, the mountainous areas of Ethiopia's southwest and South Sudan's Boma plateau are where Arabica coffee is occasionally found (Kademi et al., 2019). West and East African nations like Ghana, Guinea Bissau, Guinea, Liberia, Nigeria, Cote d'Ivoire, Cameroon, Uganda, and the southernmost country of Angola are major producers and exporters of Robusta coffee (Davis et al., 2006). Vietnam, Colombia, and Indonesia round out the top five global coffee producers (Winkler-Moser et al., 2015). Uganda, Burundi, Rwanda, and Ethiopia are the top exporters of coffee in Africa, according to DaMatta et al. (2007). The annual global production of coffee is currently around 145 million 60 kg bags, whereas the consumption is currently at 152.1 million and is rising every year (Alves et al., 2017). Brazil is the world's largest producer and exporter of coffee, as well as being the second-largest consumer (Gouvea et al., 2009). The International Coffee Organisation (Wright Jr et al., 1997) states that coffee is one of the most valuable primary products, generating US$70 billion in retail sales annually in international

trade. According to Garattini (1993), coffee contains more than 900 different components. Sugar (49%) and fats (19%) are the main components of coffee, followed by water (11,6%), proteins (11%), minerals, vitamins, and acids (7%), no nutritional components (1,5%), and free amino acids (0,8%) on wet basis. Caffeine, a variety of phytochemicals, including phenols, lactones, diterpenes, niacin, and trigonelline, a precursor to vitamin B3, as well as minerals like potassium and magnesium, are all prevalent in coffee (Cano-Marquina et al., 2013). The botanical variety, climatic factors, maturation stage, and technological process all affect the chemical makeup of coffee. The antioxidant activity is greater than commodities that are used for beverages such as cocoa, herbal tea (Richelle et al., 2001) cola and varieties of fruit juice (Pellegrini et al., 2003). Apart from its use as a beverage, it has many applications in the pharmaceutical and cosmetic industries (Aguilera et al., 2019).

The second most consumed beverage in the world (behind water) is coffee, which is also one of the most vital raw materials used in international trade. The desire to offer consumers consistently high-quality goods at a reasonable cost is what spurs interest in coffee quality evaluation. According to Barbin et al. (2014), a high-quality product is the cornerstone of success in today's intensely competitive market, so quality is undoubtedly a significant factor in the modern coffee industry. The chemical composition of coffee beans, which includes carbohydrates, proteins, lipids, polyphenols, and antioxidants, has an impact on the taste of the beverage (Ayu et al., 2020). Coffee is a consumable product, and ground roasted coffee is particularly vulnerable to adulteration because it has some physical characteristics

(such as particle size, texture, and color) that can be easily imitated by roasting and grinding a variety of biological materials, such as cereals, seeds, roots, and parchments (Reis et al., 2013). So, this food item has been the target of fraudulent admixtures with many different kinds of agricultural residues, such as twigs, coffee husks, and used coffee grounds, as well as other roasted grains like corn, barley, maize, and soybean (Oliveira et al., 2009).

Currently, Arabica coffee makes up roughly 70% of all coffee production, with Robusta making up the remaining 30% (Eira et al., 2006). The two varieties differ not only in terms of their botanic, chemical, and sensory properties but also in terms of market value, with Arabica coffee obtaining prices that are 20–25% higher than those of Robusta (Buratti et al., 2015). Currently, coffee is sold according to its varietal and/or geographic origin (Martın et al., 1999). Because Arabica beans are thought to have a finer flavor than Robusta and are therefore valued by consumers, they are more highly valued by the trade. Currently, Arabica and Robusta roasted beans or mixtures of these two varieties are used to make the majority of commercially available coffee beverages (Esteban-Dıez et al., 2004). Food processors and regulatory agencies are concerned about the possibility of deliberate or fraudulent mislabeling as a consequence (Downey & Boussion, 1996). It is therefore crucial that the different types of raw beans and coffee products can be properly distinguished. However, these physical indicators are lost during roasting and milling, necessitating the application of alternative methods to identify ground roasted coffees (Kemsley et al., 1995). Although Arabica coffee is generally regarded as superior to Robusta because of its fine and aromatic flavour, the latter

in certain regions is a viable competitor due to lower production cost, higher yield, significantly lower price (Mendes et al., 2001), and higher amount of caffeine (DaMatta et al., 2007). The authenticity of the food and its product has now become a major challenge mainly because it is related to food fraud. Analytical techniques are important for the quality and safety control of food commodities to prevent fraud. In food analysis, Near Infrared spectroscopy has been used since 1938 (Huck et al., 2005). Spectroscopic methods combined with chemometrics techniques are usually used for food traceability and authentication (Mees et al., 2018). This is due to the advantages of rapidness, high sensitivity, convenience, and simplicity.

According to Slaughter et al. (2001), Norris (1964) was the first to use NIR spectroscopy for agricultural purposes to determine the moisture content of grain. Since then, it has been used to rapidly assess a variety of agricultural and food products' moisture, protein, and fat contents (Gunasekaran & Irudayaraj, 2000). Due to its advantages over other analytical techniques, the most notable of which is its capacity to record spectra for solid and liquid food samples without prior manipulation, near-infrared spectroscopy has recently gained wide acceptance in a variety of fields. Additionally, advances in instrumentation have led to the creation of spectrophotometers that can quickly produce spectra that are adaptable enough to be used in a variety of settings; as a result, portable equipment can record spectra on-site or even at production sites (Blanco & Villarroya, 2002).

The NIR technique records spectra quickly, and the pretreatment of samples is essentially unnecessary, which speeds up analysis. The ability to conduct field measurements rather than needing to collect samples for later analysis in the lab

influences analysis speed in another way. On-site measurements are possible with some NIR spectrophotometers. Portable NIR spectrophotometer development has been accelerated by the miniaturization of optical components. Current models that use such optical devices include handheld instruments and equipment that can be mounted on vehicles like tractors or hand-held instruments and equipment that can be carried in a backpack (Rosenberg Jr et al., 2000).

The analytical information found in the typically broad, extensively overlapped band of NIR spectra is hardly selective and is affected by a variety of structural, chemical, and physical parameters. Furthermore, extremely small spectral variations that are challenging to identify with the naked eye may result from variations between samples. To obtain as much pertinent information from the analytical data as possible, chemometrics must therefore be utilized (Blanco & Villarroya, 2002; Siesler, 2007). The increasing demand for quality assurance in the food industry necessitates the use of sophisticated analytical techniques for impartial quality control. However, conventional analytical techniques are labor-intensive, expensive, and time-consuming. NIR spectroscopy provides a simple, quick, and affordable substitute. Online and in-person applications are possible because NIR spectroscopy allows measurement without prior sample preparation. As a result, this method satisfies the criteria for industrial application for continuous procedure and quality monitoring.

1.2 Problem Statement

Food adulteration is the process of adding harmful, unnecessary, or useless substances to food, which lowers the food's quality and has several negative health

effects. Food that has been altered in any manner may be toxic, deficient in nutrients necessary to maintain a healthy lifestyle, and able to trigger allergies in people with certain sensitivities (Bansal et al., 2017). Coffee is globally one of the most consumed beverages. It plays an important role in the human diet; in addition to an appetizing taste, they are rich in antioxidants and polyphenols. It is the primary source of income for 25 million people worldwide. Given that many of the nations that export coffee are developing nations, the money generated through the sector aids in the transfer of wealth from middle- and high-income countries to developing countries. About 167.26 million 60kg per bags of coffee were anticipated to be consumed worldwide in the coffee year 2020–2021, up 1.9% from the 164.13 million bags that were reported in the previous year. There has been an increase of 2.1% in the coffee year 2021/22 at estimated 170.3 million bags (ICO, 2022a). This demonstrates that not only does the demand for coffee keep on increasing, but that coffee is also becoming a more significant and financially viable export good on the global market.

Additionally, coffee export is essential for paying off foreign debt (Nestle, 2004). Coffee's purity and the identification of external impurities have been a constant concern due to its significant economic importance for the nations that produce, export, and import coffee. Roasted coffee and powder are frequently and in various ways adulterated. To reduce the price of coffee blends, it may be necessary to alter the quality of the beans (taking into account species, geographic origin, and defective beans) as well as to add additional ingredients (coffee husks and stems, maize, barley, chicory, wheat middlings, brown sugar, soybean, and rye)

to the coffee beans (Toci et al., 2016). It is crucial to be able to correctly identify the different types of raw beans, roasted and ground coffee, and different coffee products. Also, the geographical differentiation of African coffee types (bean, roasted, powdered) can be correctly recognized. However, these physical indicators are lost during roasting and milling, necessitating alternative methods for identifying ground roast coffees (Kemsley et al., 1995). The standard techniques for analyzing coffee beans are expensive, tedious, time-consuming, destructive, chemical-intensive, and call for meticulous sample preparation. Therefore, a solution to address these shortcomings must be sought after. The potential of infrared spectroscopy to offer a solution appears very promising.

1.3 Justification

Applications of infrared spectral analyses have increased in the quality assessment of food products, and they are frequently employed for predicting the characteristics of grains, tea, honey, etc. It is easy, quick, non-intrusive, non-destructive, and chemical-free, so it is also environmentally friendly. Using NIR spectroscopy, Esteban-Díez et al. (2007) were able to distinguish between arabica and Robusta as well as a blend of the different varietal compositions between 1100 and 2500 nm.

Buratti et al. (2015) classified green coffee beans, roasted coffee, and coffee beverages for both arabica and Robusta coffee using NIR spectroscopy. There are many techniques used to detect the presence of adulterants. However, no sample preparation is needed and minimal waste is produced when using NIR (Rodriguez-Saona & Allendorf, 2011). Their use has been considered to discriminate against

non-defective and substandard coffee beans (Craig et al., 2012), adulteration in coffee (Ebrahimi-Najafabadi et al., 2012), assess caffeine, theophylline, and theobromine in coffee (Huck et al., 2005), ash and lipids in coffee (Pizarro et al., 2004), the effect of roasting conditions (Lyman et al., 2003), as well as to monitor the roasting process, measure moist sucrose and colour in coffee (Santos et al., 2016). The merits are: a rapid technique that requires minimal sample preparation is accurate, environmentally friendly, non-invasive, semi/non-destructive, and allows simultaneous analysis.

NIR spectroscopy has successfully been used for the geographical differentiation of coffee beans (Giraudo et al., 2019; Marquetti et al., 2016). Other uses of near-infrared spectroscopy are the Identification of coffee leaves (Mees et al., 2018), prediction of antioxidants in roasted and spent coffee (Catelani et al., 2017a; Páscoa et al., 2013), coffee adulteration (Pizarro, Esteban-Díez, & González-Sáiz, 2007; Ryckewaert et al., 2020), coffee quality (Baqueta et al., 2021) and prediction of caffeine in coffee (Budiastra et al., 2018).

1.4 Objectives of the Study

1.4.1 General objective

This study aims to use the handheld infrared spectral technique to develop a predictive model for the identification of adulterants and rapid estimation of the quality of coffee products

1.4.1.1 Specific objectives

In order to achieve the general objective, the following were the specific objectives:

1. To authenticate different states of coffee bean varieties by using handheld NIR spectroscopy
2. To differentiate geographically African coffee types (bean, roasted, powdered) by handheld NIR spectroscopic method
3. To detect adulteration of coffee with coffee husk using a portable NIR spectrometer and chemometric technique.
4. To determine the proximate and mineral parameters of commercially sold coffee powder.
5. To assess the microbial and quality parameters of commercial coffee powder marketed in Ghana

1.5 Research Questions

1. What different states of coffee bean varieties can be authenticated by handheld NIR spectroscopy?
2. Which African coffee types (bean, roasted, powdered) can be differentiated geographically using the handheld NIR spectroscopic method?
3. How will coffee husk be detected in Robusta coffee using a portable NIR spectrometer and chemometric technique?
4. How will the proximate and mineral parameters of commercially sold coffee powder be determined?
5. How will the microbial and quality parameters affect the market of commercial coffee powder in Ghana?

1.6 Linkages

This study aims to apply the handheld infrared spectral technique to develop a predictive model for the identification of adulterants, and rapid estimation of the quality and safety of coffee products. The subsequent paragraphs demonstrate the interconnectedness of the five papers developed from the research and their integration in helping to achieve the purpose of the study.

The first paper of the study authenticates rapidly coffee bean varieties of different forms using NIR spectroscopy and multivariate data modeling. After its development, the model was used to test the different forms (raw, roasted, powdered) of the coffee varieties. The second paper explored the practicality of developing a model for authenticating African geographical coffee types by NIR. After its development, the model was used to test the different geographical coffee types (raw, roasted, powdered). The third paper focused on the application of NIR spectroscopic detection of coffee adulteration with coffee husk. The model developed was used to test adulterated coffee samples. The fourth paper evaluated some chemical quality parameters of commercially sold coffee powder. The chemical parameters were proximate and mineral parameters of the coffee. Moisture and ash content were high because of impurities in the coffee samples and also samples were not well dried. On the other hand, the fifth paper determined the biological contaminants of the samples used for the chemical analysis. The microbial analysis was thought to be crucial because some coffee producers do not use good management practices in the processing of the coffee.

1.7 Organization of the Thesis

This thesis consists of eight (8) chapters. Chapter one (1) highlights the background of the study, the main and specific objectives, and the justification of the study. Chapter two (2) gives an overview of relevant literature and theoretical foundations on the research subject. Chapters 3, 4, 5 6, and 7 are dedicated to the research articles that are specific to the objectives stated above. The summary, key findings, conclusion and recommendations are presented in Chapter 8.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

Coffee is a tree crop that flourishes best in the tropical highlands and lowlands areas of countries that produce coffee. The Arabic word "quahweh," which according to some was originally a poetic term for wine, is where the word "coffee" comes from. With an annual production of about nine million tons of green beans, it is a significant agricultural product produced in about 80 tropical countries, supporting an estimated 125 million people in Latin America, Africa, and Asia (Krishnan, 2017). The genus Coffee consists of 103 species, all of which are found only in the tropical forests of Africa, Madagascar, and the island of the Indian Ocean (Zerga & Tsegaye, 2020). Only two of the species, Coffea arabica L. and Coffee canephora Pierre ex Froehn, are valued commercially in the global coffee market. The coffee plant, coffee arabica, is native to Ethiopia. Around AD 850, the Arabian colony of Harar began cultivating it there as well. Later, coffee varieties like Robusta and Liberia were found in Africa (Smith, 1985). The physical characteristics of these coffee beans, such as their size, shape, color, and chemical composition (such as the presence of minerals, caffeine, chlorogenic acid, volatile compounds, and higher molecular weight materials), are what characterize them (Nunes & Coimbra, 2001; Rubayiza & Meurens, 2005).

Millions of people, especially in developing nations, depend on coffee for their livelihoods. It is produced in the majority of African nations and is crucial to

the economies of Ethiopia, Cote d'Ivoire, Uganda, Zimbabwe, the Democratic Republic of the Congo, Angola, Ghana, Rwanda, Tanzania, Cameroon, and other nations (ICO, 2009b). In the first quarter of 2017, the world's production surpassed 159.663 million 60 kg per bag, with Brazil accounting for 40% of the total. The main commercialized species, coffee Arabica (59%) and coffee Robusta (41%), contributed to the nearly 172 million 60kg per bag of coffee produced globally in 2020–21. Brazil is the world's largest producer and exporter of coffee, with a projected total production of 69 million 60kg per bag in the crop year 2020–2021. The European Union, the United States, and Japan are three of the top importers of coffee from Brazil (ICO, 2018).

2.2 Anatomy of Coffee Cherry

The coffee seed is composed of an egg-shaped, convex plane surface characterized by a longitudinal furrow. The husk, or exocarp, of the mature coffee fruit, has a color that varies from yellow to red depending on the genotype of the species and has a hue of green when the fruit is immature. The fleshy, readily removable mesocarp is made primarily of pectin, fructose, and other sugars like glucose. According to Janissen and Huynh (2018), the pectin layer sometimes referred to as mucilage, contains proteins, fat, lipid minerals, tannins, polyphenols, and caffeine. Endocarp, also known as parchment, is a polysaccharide layer that resembles a thin, yellowish, and easily shreddable piece of paper made primarily of lignocellulose (Esquivel & Jimenez, 2012). The integument, also known as silver skin, is the layer that surrounds the bean and is made up of monosaccharides, proteins, polyphenols, and phenolic compounds with significant antioxidant

activity, as well as polysaccharides like cellulose and hemicelluloses (Farah & dos Santos, 2015; Janissen & Huynh, 2018). The two hemispheres of elliptical seeds that collectively make up a coffee bean contain endosperm and embryos (Esquivel & Jimenez, 2012; Farah & dos Santos, 2015). The Arabica and Robusta coffee-growing regions had a different impact on the weight, size, and volume of the beans. The characteristics of the Robusta and Arabica parchment were observed to differ, with the latter being larger and having a generally elongated shape both longitudinally and laterally in comparison to the former variety (Chandrasekar & Viswanathan, 1999).

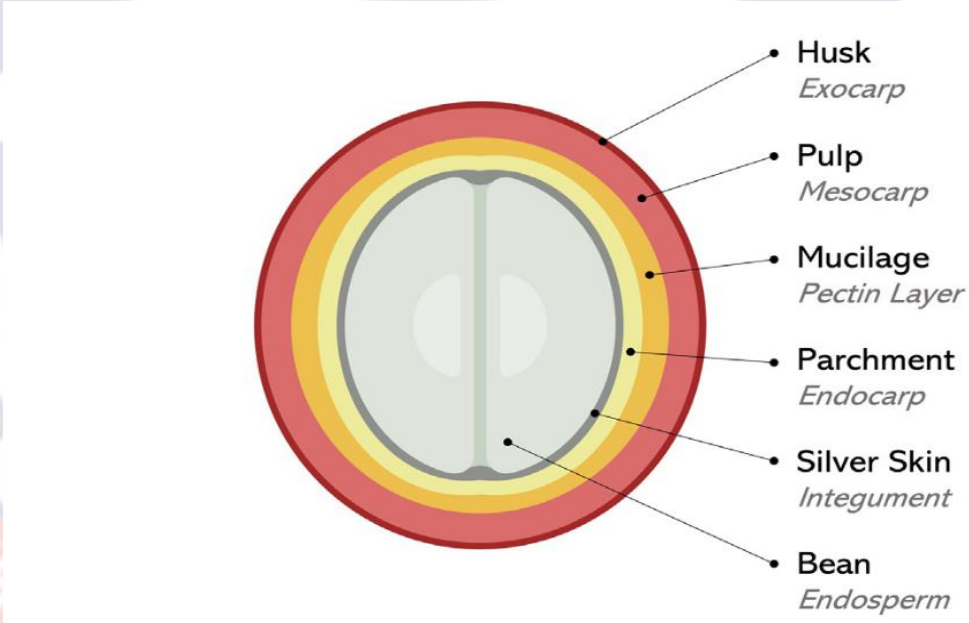


Figure 2.1: The anatomy of the coffee cherry (de Melo Pereira et al., 2020)

2.3 The by-products of coffee

According to Esquivel and Jimenez (2012), the main by-product of the dry method is the coffee husk, which is made up of pulp, dried skin, and parchment, while the by-product of the wet method is pulp and coffee silver skin. Coffee extract residue or used coffee grounds are other by-products that are produced during the

brewing process (Murthy & Naidu, 2012). To obtain coffee beans, various processes must be used, and many by-products are produced (Mussatto et al., 2011; Nabais et al., 2008), of which about 50% are not used to produce green coffee beans. As a result of the by-products' high concentration of caffeine, tannins, and polyphenols, they are highly polluting and cannot be used as animal feed in larger quantities (Bondesson, 2015).

2.4 Coffee Production and Consumption

More than 400 million cups of coffee are consumed worldwide each year, making it one of the oldest cultures still in existence. Blends of roasted and ground Arabica and Robusta coffee beans can be used to make coffee beverages (Buratti et al., 2015; Martini et al., 2016). From cultivating 11 million hectares of coffee, production increased from approximately 8.5 million tons in 2008 to 10.7 million tons in 2020 (FAO, 2022). According to Bermudez et al. (2022), approximately 80% of coffee production was exported in 2021–2022, up from 64% in 2020 and 74% in 2019. This represents a significant source of foreign exchange earnings for exporting nations. Since 2016, Brazil, Vietnam, and Colombia have been the top three exporters and countries with the highest production rates, with respective exports of 33 million, 29 million, and 14 million 60 kg per bag in 2021–2022. As of 2021/2022, the top three importers were the European Union (UE), the United States, and Japan, importing 43 million, 26 million, and 7 million 60 kg per bag of coffee respectively (Bermudez et al., 2022).

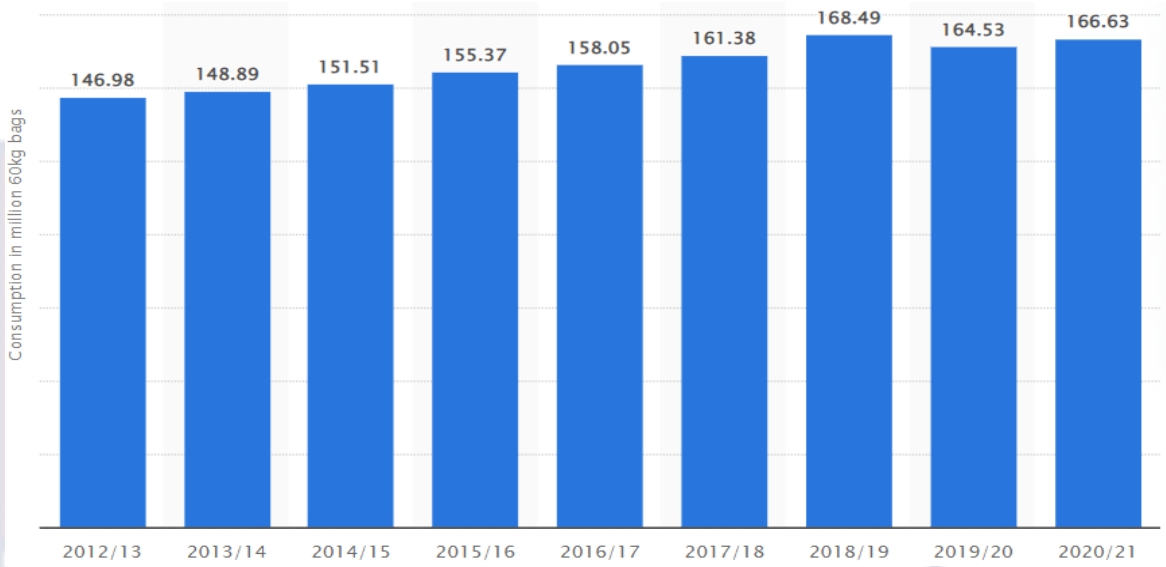


Figure 2.2: World coffee consumption for the past five years in million 60 kg bags (Conway, 2020)

2.5 Varieties of Coffee

2.5.1 Robusta coffee

Robusta plants can be grown at low altitudes between 200 and 800 meters and are more hardy, higher-yielding, and adaptable (Baker, 2016). In comparison with arabica coffee beans, Robusta coffee has higher levels of caffeine (Briandet et al., 1996), chlorogenic acid (Moon & Shibamoto, 2009), and antioxidants (Budryn & Nebesny, 2008; Daglia et al., 2000; Sacchetti et al., 2009). Coffee Robusta, grown on native smallholdings in West Africa, is renowned for its disease resistance, thus its specific epithet.

2.5.2 Arabica coffee

Arabica plants are delicate and grown between 600 and 2000 meters above sea level (Baker, 2016). Arabica is the most expensive because it is typically valued more for its organoleptic characteristics (Briandet et al., 1996). Due to its reputation

for having a more refined flavor than Robusta and consequently higher consumer appreciation, it is more highly valued by the trade and the coffee that is currently most widely sold. In comparison to the Robusta coffee bean, Arabica coffee exhibits a more pleasant, more distinctive, and less bitter flavor.

2.6 Coffee Quality

The desire to offer consumers consistently high-quality goods at a reasonable cost is what promotes interest in coffee quality assessment. Since a high-quality product is a prerequisite for success in today's extremely competitive market, quality is undoubtedly a significant factor in the modern coffee industry. Bean size, color, shape, roast potential, processing method, crop year, flavor or cup quality, and presence of defects are among the factors frequently used to assess the quality of coffee beans (Franca & Oliveira, 2008). Generally speaking, Arabica coffee is thought to be of higher quality, and as a result, it costs more than Robusta. The adulteration of roasted coffee is also an approach used to reduce costs.

Consequently, the coffee industry is becoming more interested in defining its production (i) because of the widespread production and consumption of coffee, (ii) because it has positive nutritional effects on human health, and (iii) because it is important to prevent adulteration of coffee given the wide range in final selling prices depending on the type or region of origin of the bean. Given the significant price variations at the point of sale, assurance of high-quality roasted coffee has drawn a lot of attention to reduce and eliminate coffee adulteration (Pizarro, Esteban-Díez, González-Sáiz, et al., 2007).

2.7 Authentication of Food

According to Esteki et al. (2018), food authentication is the process by which food is examined for quality, safety, and compliance with its label description, consumer protection laws, and applicable standards. Food authentication is a significant issue that has gained importance recently. The food industry, regulatory agencies, and consumers are all interested in authenticating raw materials and food products to satisfy food quality and safety requirements. However, proving the true origin of a product and determining whether it has been mixed with adulterants is a major challenge for all three groups (Pizarro, Esteban-Díez, González-Sáiz, et al., 2007). There is a great deal of concern about the highly destructive effects of food falsification on the global food supply chain due to the growing number of food products available and incidents of adulteration causing serious economic and health losses.

Food imports and exports from many nations have increased as a result of the globalization of markets, to the point where traceability in the food production chain is being neglected. Since the behavior of farmers, producers, manufacturers, retailers, and consumers involved in their production, distribution, exchange, and consumption affects the quality of final products, food authenticity has consequently become a major concern for the food industry. Labeling laws obliging them to list the main components of each product are binding on food manufacturers and producers. Customers value quality indicators that are simple to understand, like certified quality labels, geographic cues, or guarantee seals, as well as required minimum requirements like best-before dates. Food authentication

procedures are designed to find imitations, prevent unfair competition in the common market, and safeguard consumers from deception.

2.8 Main Problems of Quality and Safety of Coffee

2.8.1 Adulteration of food

As long as there has been food production, there has been the possibility of adulteration of food, which involves the addition of inferior ingredients or the removal of essential food components. For consumers to choose food products for consumption, there must be accurate, truthful information available. Frequently, lifestyle and health considerations influence consumers' purchasing decisions. Economically, adulteration is now recognized as posing a serious risk to the public's health due to growing consumer awareness of the issue. According to Scarano and Rao (2014), among other things, food products can be tainted by adding illegal substances, substituting a component with a comparable but less expensive substitute, artificially extending shelf lives, claiming false or on-applied processes, making false claims about the number of ingredients, and disclosing false information about the origin of the product's production.

The commercial value of arabica has made it the subject of countless and increasingly sophisticated adulterations over time (Arrieta et al., 2019), primarily through the addition of roasted barley, corn, rice, and coffee husk. Robusta coffee is frequently used to adulterate arabica coffee due to its lower market share and compositional similarity (Dankowska et al., 2017). The main coffee adulterants are roasted and unroasted coffee husk, twigs, barley, chicory, malt, starch, corn, glucose syrups, caramelized sugar, and mixing two species (adding less expensive

Robusta to pure arabica coffee) or mixing expensive coffee beans from one growing region with inexpensive beans grown in another region (Prodolliet et al., 1995). This method of production affects the sensory qualities and quality of coffee beverages (Esquivel & Jimenez, 2012).

2.8.2 Microbial contamination of coffee

Coffee cherries and beans, like other crops, are subjected to contamination and microbial colonization as a result of various stages of plant growth, harvesting, transportation, and storage. According to Batista et al. (2003), microbial action has a significant impact on environmental factors, crop and product handling, as well as the quality and safety of the final product. According to microbiological research, toxigenic fungi genera like Aspergillus and Penicillium are inherent contaminants of coffee cherries and beans that exist from the farm to the warehouse (Silva et al., 2008). They cause significant damage to coffee beans and may be bad for the coffee's organoleptic and sanitary quality. *Aspergillus* and *Penicillium* produce ochratoxin A (OTA) to contaminate coffee beans when present. One of the most prevalent mycotoxins found in coffee beans and roasted coffee is ochratoxin A, which has been linked to several harmful health outcomes, including hepatotoxicity, teratogenicity, and carcinogenicity (Clark & Snedeker, 2006; Joint & Additives, 2001).

2.8.3 Acrylamide formation

Thermal processing of food, whether industrial or domestic, has long been used to enhance its organoleptic qualities, microbiological safety, and preservation

(Studer et al., 2004). Thermal food processing may also result in the formation of heat-induced toxic food contaminants, including acrylamide and furan, which have genotoxic and carcinogenic effects (Morales et al., 2009; Tritscher, 2004). The Maillard reaction, a series of non-enzymatic reactions between reducing sugars like glucose and fructose and free amino acids, primarily asparagine, produces acrylamide, a thermal processing contaminant (Becalski et al., 2003; Mottram et al., 2002). Roasted coffee has been shown to contain high levels of acrylamide (Batista et al., 2003), making it a significant dietary source of acrylamide exposure with levels that may exceed the values advised by the European Commission (Mesias et al., 2022).

2.9 Roasting of Coffee

Coffee beans are heated to temperatures between 170 and 240 °C for a set period, usually not more than 15 minutes. The process can be broken down into three stages: drying, roasting, and cooling. During those phases, physical and chemical changes occur, and the reaction pathway produces a variety of volatile compounds that are quite prevalent in coffee headspace (Gonzalez-Rios et al., 2007). The chemical composition of the roasted beans determines how good a cup of coffee becomes. During all stages of coffee roasting, a variety of different chemical compounds contained in the raw coffee beans interact and react with one another to produce vastly different final products (Ribeiro et al., 2011). During roasting, some elements of the coffee, including water, volatile flavors, chemicals, and water, escape from the coffee beans and contribute to the loss of coffee bean mass. According to a previous study, the water content of roasted coffee beans

affects how the beans grind, how much coffee is extracted, and how long the aroma lasts (Schenker et al., 2002). Grinding of coffees with higher water contents resulted in less fine particles, and hence faster percolation times. In addition to being highly sensitive to water content during open and closed storage, the evolution of hexanal and sulphides, which contribute distinctive flavours from roasted coffee beans was also observed (Baggenstoss et al., 2008).

In the coffee industry, the same or different coffee blends are typically roasted at various intensities to produce coffees that have a variety of qualities and are marketed under various brands. The technological processes of roasting and grinding have an impact on the properties of the finished product in addition to environmental factors and the harvesting process (Bressanello et al., 2017). The degree of roast largely determines each type of coffee's essential qualities, including taste, aroma, and purity. According to Correia et al. (2018), the procedure entails gradually raising the temperature while the raw coffee bean is being processed. This lowers the moisture content of the bean to 3wt% (weight of solute/weight of solvent*100)

2.10 Characterization of Coffee

Chemical characteristics of coffee include its metal content, volatile components in roasted and beverage coffee, chlorogenic acids, caffeine, trigonelline, aqueous extract, amino acids, and polyphenols in green coffee, as well as the fatty acid profiles of ground coffee and the tocopherol profiles of Arabica and Robusta, both green and roasted (Tugnolo et al., 2019).

2.10.1 The biochemical component in coffee

2.10.1.1 Antioxidants in coffee

Coffee has considerably greater antioxidant activity than other drinks like cocoa, green tea, black tea, herbal (Richelle et al., 2001), cola, beer, and different kinds of fruit juice (Pellegrini et al., 2003). Coffee is regarded as the primary beverage source of phenolic antioxidants (Mattila et al., 2006). According to studies, coffee beans that were roasted under light to medium conditions typically produced the highest levels of antioxidant activity (Cämmerer & Kroh, 2006). Aroma, chemicals, in particular volatile heterocyclic compounds, formed in roasted coffee have been reported to have antioxidant properties (Yanagimoto et al., 2002). Antioxidant supplementation has gained a lot of attention recently as a method of preventing various diseases (cancer, cardiovascular, Parkinson's) brought on by oxidative damage. Recently, coffee has drawn a lot of attention as a beneficial source of supplemental antioxidants.

2.10.1.2 Polyphenols

Plants produce large amounts of polyphenols as secondary metabolites. The primary polyphenols in coffee are chlorogenic acids, which are glycosylated derivative forms of the compound (Manach et al., 2004). According to Fraga et al. (2010), there are several biological activities associated with polyphenols, such as flavonoids, that are advantageous to human health. Dietary polyphenols, which are found in fruits, vegetables, cereals, legumes, and beverages (tea, coffee, and wine), are known to prevent several diseases, including osteoporosis, cancer, cardiovascular disease, and diabetes (Scalbert et

al., 2005). These polyphenolic biological activities were closely linked to their antioxidant activity (Fraga et al., 2010).

2.10.1.3 Proximate composition

The composition of commonly consumed foods in quantities that supply macro- and micronutrients is important for the overall evaluation of the state of the public's health. Analysis of foods will offer proof of their nutritional value, serve as a guide for making healthy choices, and encourage the consumption of varieties with superior qualities in times of illness and the prevention of disorders linked to diet. An important index for evaluating the nutritional potential of crops is the determination of the proximate composition of food. The chemical makeup of the roasted coffee bean, which is influenced by the chemical composition of the green beans and the post-harvest processing conditions, determines the quality of coffee used for beverages (Franca & Oliveira, 2008).

According to Clarke (2012), variations in species/variety, as well as other factors like origin, agricultural practices, growth, storage situation, and degree of maturation, are all related to variations in the proximate composition of green and roasted coffee. Due to pyrolysis reactions, roasting causes significant changes in chemical composition. The proximate composition, however, essentially remains the same because of changes within a particular class of compounds.

2.10.1.4 Mineral composition of coffee

It has been discovered that green, roasted, ground and coffee infusions contain about 30 different elements. Major elements, minerals or micronutrients (such as Ca, K, Mg, Na, S, and P), minor elements or micronutrients (such as Cl, Co, Cr, Cu, Fe, and Mn),

and trace elements (such as Al, As, B, Ba, Cd, and Hg) can all be divided into these three categories. According to Pohl et al. (2013), the chemical composition of roasted and ground coffee is generally influenced by the region where the coffee beans were grown, which is primarily determined by the soil's characteristics, the type of coffee grown, and how the plants were grown. When compared to unprocessed coffee of the same variety, roasted coffee has higher concentrations of K, Na, Ca, Mg, and Fe (Amorim Filho et al., 2007). Due to their revealed nutritional and dietary value, minerals are necessary, especially in light of the significant global coffee consumption (Oliveira et al., 2012).

2.11 Health Effects of Coffee

Coffee is one of the most widely consumed beverages in the world, but due to the stimulant effects of caffeine on the central nervous system and its adverse effects on the cardiovascular, central nervous, and endocrine systems (Panagiotakos et al., 2003; Rasch, 2003; Virtanen et al., 1994), consumption has significantly decreased. On the other hand, research has shown that modern society can benefit from the preventive effects of caffeine and coffee, which both have a stimulant effect. According to reports in the literature, drinking coffee daily in moderation has been shown to improve the clinical conditions of diabetic patients and lower their risk of developing type 2 diabetes, increase their energy expenditure, and reduce their risk of weight gain (De Matteis et al., 2002; Ryan, 1996). Coffee is a functional food with antioxidant properties that lower the risk of mortality and the incidence of liver disease, cancer, and Parkinson's disease (Dórea & da Costa, 2005).

2.12 The Method used for Differentiating Coffee Varieties and Adulterant

A simple visual inspection is an appropriate method for differentiating between genuine coffee samples and fraudulent ones, and several analytical strategies have been developed. Also, the use of chemical analyses is crucial to characterize coffee matrices, but they can be time-consuming, expensive and needs expert laboratory staff and complex instrumentation.

2.12.1 Analytical methods

Analytical techniques are essential for the safety and quality control of food products or supplements based on natural compounds and to prevent fraud. The following are the analytical methods used to check the quality of coffee; Micro-Raman spectroscopy, HPLC analysis, Solid phase microextraction (SPME), GC-MS analysis, Vibrational spectroscopy, and others.

2.12.1.1 Vibrational spectroscopy

The near-infrared (NIR, 14000 to 4000 cm^{-1}), mid-infrared (MIR, 4000 to 400 cm^{-1}), and far-infrared (400 cm^{-1}) wavelength ranges are all acceptable infrared spectroscopy methods. Only the MIR and NIR regions are usually used for food authentication, but because infrared spectroscopy is simple to use and requires little setup, it has gained popularity as a versatile research tool (Abbas et al., 2020; Rodriguez-Saona & Allendorf, 2011). When performing routine authentication analyses, infrared spectroscopic methods offer easy-to-use, quick, and affordable tools for preliminary sample screening.

2.12.1.2 Near-infrared spectroscopy (NIR)

According to Huck et al. (2005) and Pasquini (2003), near-infrared spectroscopy, which relies on molecular overtones and combination vibrations, is extremely effective for probing bulk materials with little to no sample preparation. When used in-line for quality control during industrial coffee processing, it is a quick substitute for traditional methods. The coffee industry has recently shown an increasing interest in the creation of calibration models using chemometric methods in conjunction with various techniques for quantitative and qualitative analyses of green and roasted coffee (Barbin et al., 2014).

The use of near-infrared (NIR) spectroscopy in this area has grown as a result of its speed, simplicity, and safety as well as the ability to determine multiple parameters at once with little to no sample preparation (Marquetti et al., 2016). Due to its speed, accuracy, and capacity to provide spectra for both solid and liquid samples without the need for prior manipulation, it is particularly promising in this regard (Blanco & Villarroya, 2002). This method is relatively inexpensive and suitable for implementation in the daily operations of industries for a wide range of applications because it requires little to no sample preparation (Porep et al., 2015). The most popular method for authentication and traceability is the use of spectroscopy in conjunction with chemometric techniques (Nescatelli et al., 2017). Some of the spectrometers are portable and can be easily moved from one place to another to permit in-field and product-to-product evaluation.

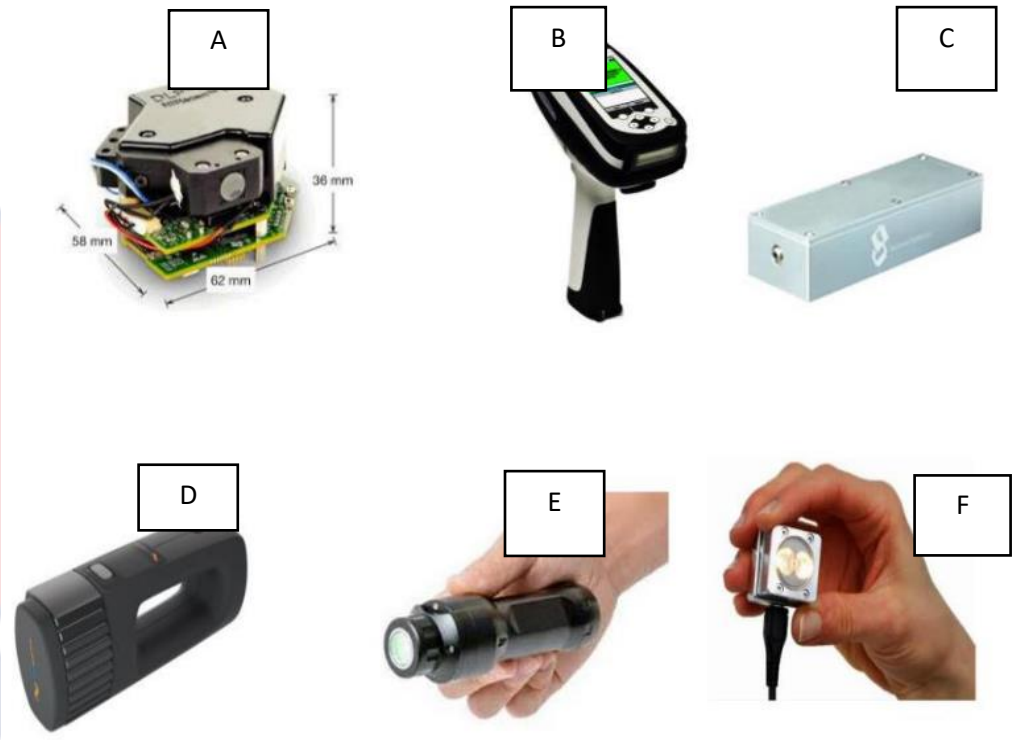


Figure 2.3: Selected portable NIR spectrometers (Beć et al., 2021)

A) DLP NIR scan Nano evaluation module (EVM; Texas Instruments);

B) MicroPHAZIR (Thermo Fisher Scientific); C) nano FTIR NIR (SouthNest Technology); D) NeoSpectra (Si-Ware Systems); E) MicroNIR Pro ES 1700 (VIAVI); and F) NIRONE Sensor S (Spectra Engines).

2.13 Procedure for NIR Spectroscopy and Chemometrics

2.13.1 Spectral acquisition

Scanning to obtain spectra and using chemometrics to extract meaning from the spectral data are the general steps typically used for qualitative and quantitative measurement using NIR spectroscopy.

2.13.2 Chemometrics

Chemometrics is the science of using mathematical or statistical methods to relate measurements made on chemical systems or processes to the state of the system (Cen & He, 2007). Numerous studies have shown that the spectra of samples obtained by spectrometers, either on whole or semi-destructive coffee beans and coffee products, are accurate. Pretreatment/preprocessing, model development or building, model testing and transfer are the four main steps. Chemometrics was utilized throughout the entire process to analyze NIR spectroscopy and develop the final method.

2.13.3 Pretreatment

The pre-processing methods that have been employed by several researchers are first derivative (FD), second derivative (SD), mean centering (MC), multiplicative scatter correction (MSC), and standard normal variant (SNV).

FD. - Baseline offset can be eliminated with great efficiency using FD pre-processing. One of the best ways to eliminate baseline weaknesses is through spectral derivative transformation. (Duckworth, 2004).

SD. – The SD pre-processing technique enhances small spectral differences while removing baseline drifts, the linear trend from a spectrum, and detached overlapping peaks (Duckworth, 2004; Rinnan et al., 2009). The use of derivative chemometric tools facilitates the detection of minute chemical variations or differences between the studied samples' spectra.

MC. – The MC pretreatment method is accomplished by deducting the calculated average spectrum of the data set from the average values of the spectrum. This was

carried out by Teye and colleagues who used this as a pretreatment technique to ensure interpretable results (Teye et al., 2013).

MSC. – MSC is a unique pretreatment technique that is used to correct scattered light on various particle sizes. The basic idea is to correct spectral effects that are additive and multiplicative. It is also helpful for minimizing or eliminating unhelpful scattering-related variability, which lessens the effects of light emission and the spectrum's diversity. MSC assumes that every sample has the same level of scatter as a reference spectrum. Using the findings of simple linear regression estimation, the procedure entails studying each spectrum and correcting it per the reference spectrum (Dhanoa et al., 1994). Additionally, it stops scattered light from various particle sizes.

SNV. - The SNV transformation is applied to each spectrum by dividing by the standard deviation of the spectrum and subtracting its mean (Candolfi et al., 1999). The multiplicative interferences of light scatter, particle size, and altered light distance are corrected by SNV. To eliminate slope variation on an individual spectrum basis, each object experienced an independent transformation. SVN corrects scatter effects that are both additive and multiplicative (Dhanoa et al., 1994).

2.14 Multivariate Data Analyses

Multivariate techniques are grouped as qualitative and quantitative models.

2.14.1 Qualitative models

2.14.1.1 Linear Discriminant Analysis (LDA)

For data classification and dimensionality reduction, LDA is the method most frequently used. When the within-class frequencies are different, linear discriminant analysis can handle the situation with ease (Thapngam et al., 2012). This is because the performance of the algorithms has been examined using test data that was generated randomly. The maximum separability is produced by this method by maximizing the ratio of between-class variance to with-class variance in any given data set. The distribution of the feature samples can be better understood using this technique (Aziz et al., 2014).

2.14.1.2 Support Vector Machines (SVM)

SVM belongs to the group of supervised machine learning algorithms (Jiao et al., 2017). As a kernel learning algorithm, SVM is the most popular choice. The kernel method achieves nonlinear separation in the kernel space, which allows the SVM algorithm to find hyperplanes in the kernel space (Liu et al., 2015). It does this by converting the data from the feature space into higher dimensional kernel space. The Gaussian radial basis function (RBF) and polynomials are frequently used in the SVM implementation as kernel functions.

2.14.1.3 Partial least square discriminate analysis (PLS-DA)

PLS-DA is a supervised multivariant statistical tool that can compress data and extract characteristic information. It integrates the fundamental principal component analysis (PCA), canonical correlation analysis, and multiple regression

analysis functions (Tang et al., 2014). Due to its ability to manage multilinear and correlated variables through dimensional reduction, it has attained enormous popularity and wide acceptance in the field of applied research (Lee & Jemain, 2019). It is a flexible algorithm that can be used for discriminative variable selection, descriptive and predictive modeling, and both.

According to Lee et al. (2018), it has demonstrated strong performance when modeling large, multidimensional datasets for a variety of uses. It can be adapted for feature selection as well as for classification and can achieve dimensionality reduction while fully cognizant of the class labels (Christin et al., 2013; Tang et al., 2014). Additionally, it works well to choose (signal) features when classes have a clustered distribution, even if the features are buried among a lot of noise features. Additionally, it can choose the right hyperplane even with a small sample size and when there is little distance between the clusters. In contrast, PCA requires an unreasonable quantity of samples and extremely distant clusters to function.

2.14.1.4 K-nearest neighbour (KNN)

KNN classifier is a case-based learning algorithm that uses a distance or similarity function for various pairs of observations. It is referred to as a lazy learning algorithm because computations are postponed until classification and functions are only approximated locally. Because of its efficiency, non-parametric nature, and simplicity of implementation, it has been successfully applied in a variety of fields, including the medical and food industries. Finding the ideal value of K is challenging due to the lengthy classification process. A higher K in

classification reduces the impact of noise. Traditional KNN should be modified with different K for different classes rather than a fixed value for the classes to avoid this drawback (Khamar, 2013). PCs were utilized in this study as input data for the KNN model, which helped determine the model's effectiveness along with the K parameter. Simultaneously maximizing the variables PCs and K produced a very good model.

2.14.1.5 Neural network (NN)

A neural network (NN) is made up of countless interconnected neurons, each of which functions as a simple processor to perform a weighted sum of all inputs coming from either inside or outside the network. According to Thibault et al. (1990), the weighted sum is typically passed through a nonlinear transfer function (hyperbolic tangent, sigmoid, sinusoid, or threshold) to obtain the output, 0, of the neuron. Backpropagation, which involves redistributing the output errors to the network by appropriately modifying the weight matrices, is the most flexible learning algorithm for this feed-forward layered network (Thibault et al., 1990).

2.14.1.6 Random Forest (RF)

A variety of randomised decision trees are combined with an averaged aggregate of their predictions in the general-purpose classification and regression technique known as the RF algorithm. In situations where there are fewer observations than variables, it has performed exceptionally well (Biau & Scornet, 2016). Due to its straightforward turning and training parameters, which may fit

nonlinear models and yield excellent results (Breiman, 2001; Cao et al., 2012). This machine-learning algorithm is very popular.

2.14.2 Quantitative models

Some of the quantitative models are partial least square (PLS), interval partial least square (iPLS), backward interval partial least square (biPLS), synergy interval partial least square (SiPLS), genetic algorithm partial least square (GaPLS) and successive projections algorithm (SPA-PLS).

2.14.2.1 Partial least square (PLS)

PLS is a broad category of techniques for utilizing latent variables to model relationships between a set of observed variables. Along with dimension reduction techniques and modeling tools, it includes regression and classification tasks. All PLS methods start with the assumption that the observed data is produced by a system or process that is fueled by a relatively small number of latent (not directly observed or measured) variables. By using PLS, Herman Wold and colleagues were able to project the observed data to its latent structure (Rosipal & Krämer, 2006). In the field of chemometrics, PLS has drawn a lot of interest. This algorithm is now frequently used to process a variety of chemical data processing issues. Numerous applications of PLS in other scientific fields, such as food research, medicine, pharmacology, and social sciences, were made possible by the success of PLS in chemometrics (Baptistao et al., 2011).

2.14.2.2 Interval partial least square (iPLS)

According to Norgaard et al. (2000), interval PLS calculates local PLS models equidistant from the full spectrum region. Its main benefit is that it gives a

comprehensive graphical representation of the variation in X that is pertinent to the dependent y-variable. The predictive performance of all local models as well as the global full spectrum model are compared when iPLS models are developed on spectral subintervals of equal width. This has been employed in the selection of wavelengths (Leardi & Nørgaard, 2004).

2.14.2.3 Backward interval partial least square (biPLS)

They are a widely used modeling technique that is based on the choice of the wavelength variable. This method typically has high prediction accuracy, but it exhibits strong greedy search characteristics, which make the intervals chosen insufficient to reveal analyte information (Qu et al., 2016).

2.14.2.4 Synergy interval partial least square (SiPLS)

Nørgaard (2005) proposed the SiPLS algorithm. The spectral set can be divided into any number of intervals (10 to 25), and all PLS model combinations for two, three, or four intervals can be calculated (Norgaard et al., 2000). This model made it possible to reduce the computational load (Wang et al., 2013).

2.14.2.5 Genetic algorithm partial least square (GaPLS)

The most popular method for analyzing NIR spectroscopic data sets, GA-PLS, combines the benefits of GA and PLS. The PLS procedure could be integrated into the objective function derived from the optimization, and the GA could find the best values for a variety of disparate variables related to the calibration models. Due to the wavelength selection in the PLS calibration using a genetic algorithm, the GA-PLS exhibits superiority over other applied multivariate methods and

additionally offers helpful information about the chemical system (Xiaobo et al., 2010). One drawback of GaPLS is related to the fact that the search domain expands in line with the measurement of spectral intensities at a very large number of wavelengths, making it harder to identify the pertinent regions (Leardi & Nørgaard, 2004).

2.14.2.6 Successive projections algorithm (SPA-PLS)

The choice of variables in SPA is modeled as a constrained combinatorial optimization problem. Because the search is limited to a smaller set of variable subsets that are created using a series of projection operations on the matrix of instrumental response, the optimization is said to be constrained. To reduce redundancy and ill-conditioning issues, the operations are used to select subsets of variables with a low degree of multi-collinearity (Soares et al., 2006).

2.15 Application of NIR Spectroscopy

Table 2.1: Application of NIR spectroscopy technology in the coffee bean and coffee bean products

Application	Task	Wavelengths range	Chemometrics		Author(s) year
			Preprocessing	Algorithms	
adulteration	Identification of adulterants	908nm- 1676nm	PCA, 1-Der with Savitzky-Golay smoothing	PLS	(Correia et al., 2018)
	Identification and quantification of the addition of barley	4,000 – 10,000cm-1	Auto-scaling (AS)	GA-PLS	(Ebrahimi-Najafabadi et al., 2012)
	Prediction of food adulteration	4,000 – 10,000cm-1	Autoscaling (AS) and Standard normal variate (SNV)	Convolutional neural network (CNN)	(Chakravartula et al., 2022)
	Detection of coffee adulteration	400 – 2500nm	Savitzky-Golay derivative (first derivative) using polynomial order 2 and 3-point smoothing	PLSR	(Winkler-Moser et al., 2015)
Classification	Identifying the quality standard of special agroforestry coffee	937- 1655nm	PCA and HCA	DD -SIMCA	(Manuel et al., 2022)
	Coffee varietal differentiation	1100 – 2500nm	DOSC	PCA	(Esteban-Díez et al., 2007)
	Classification of Brazilian Coffee	800 – 2500nm	1-Der with Savitzky-Golay with 9 smoothing points, MSC	PCA, SIMCA, PLS-DA	(Santos et al., 2012)
	Determination of the geographical	12500- 3600cm-^{1}	SNV, MSC, 1st and 2nd Der using Savitzky-Golay	PLS-DA	(Giraudo et al., 2019)

	origin of green coffee				
	Cup profile determination in coffee blends	906- 1676nm	Reflectance spectra transformation, MSC, and 1-Der with Savitzky-Golay smoothing	PLS-DA	(Baqueta et al., 2021)
	Classification of green coffee	1200-2500nm	MSC, 2nd Der	SIMCA	(Okubo & Kurata, 2019)
Degree of roasting	Prediction of roasting temperature	4000-12,000cm^{-1}	Vectorial normalization, 1st Der, 2nd Der, or combination of the above options	PLS	(Alessandrini et al., 2008)
	Prediction of roasting colour	1100 – 2500nm	PLS, IPW-PLS, ISE-PLS, UVE-PLS	SELECT- OLS	(Pizarro, Esteban-Díez, González-Sáiz, et al., 2007)
	Monitoring and Predicting coffee roasting	1350 -2500nm	1-Der with Savitzky-Golay	PLS	(Yergenson & Aston, 2020)
	Assessing several traits of roasted coffee	900-1700nm and 1300-2500nm	PCA	PLS	(Pasquini, 2003)
Chemical Composition of Coffee	Determination of lipids and proteins in green coffee	12000 – 4000cm^{-1}	SG, SNV, MSC, 1ST and 2nd Der with Savitzky-Golay	PLS	(Zhu, Long, Chen, et al., 2021)
	Evaluation of chemical properties of green coffee	4000 – 12500cm-1	MC, MSC, SNV, 1ST and 2nd Der with Savitzky-Golay	PLS	(Levate Macedo et al., 2021)
	Prediction of moisture content in green coffee	1000 – 2500nm	SG derivatives (1st, 2nd and 3rd), SVN, OSC, MSC and EMSC	PLSR, MLR	(Adnan et al., 2017)

Table 2.2: Application of NIR spectroscopy technology in other foods and products

Food	Task	Wavelength	Chemometrics		Author (Year)
			Preprocessing	Algorithm	
Cocoa	Differentiation of Ghana Cocoa	4000 – 10,000cm^{-1}	MC, MSC, DC, 2nd -Der	LDA, KNN, BP-ANN, SVM	(Teye et al., 2013)
Tea	Estimation of taste quality and taste-related compound content in black tea	12,500 – 4000cm^{-1}	SVN	SI-PLS, BP-AdaBoost	(Chen et al., 2018)
Oils	Detection of Palm oil adulteration	900 – 1700nm	MSC	LDA, SVM	(Teye et al., 2019)
	Distinction of Extra Olive oil from other Grades	908 -1676nm	2nd Der, AS	OCKNN, OCSVM, SIMCA	(Yan et al., 2019)
Eggs	Determination of egg storage time	740- 1040nm	Savitsky-Golay with 3rd Der	ANN	(Coronel-Reyes et al., 2018)
Grains	Quantitative determination of corn	4003.2 – 9090.9 cm^{-1}	SNV, MSC	MB-PLS	(Jing et al., 2010)
Honey	Determination of adulterants in Chinese honey	6000 -10,000cm^{-1}	MC, MSC, 1st and 2nd Der	DPLS	(L. Chen et al., 2011)
Beverages	Determination of alcohol content in beverages	700 – 1100nm	1st Der with Savitsky-Golay	PLS	(Barboza & Poppi, 2003)
Dairy Products	Determination of fatty acids in milk	2500 -1000nm	MSC, SNV	PLS	(Stefanov et al., 2013)
Meat	Chemical composition of meat	400 - 2500nm	SNV, Detrend	PLS	(Prevolnik et al., 2010)
Fruits and vegetables	Evaluation of strawberry fruits quality parameters	10000 – 40000cm^{-1}	1st Der with Savitsky-Golay	PLS	(Mancini et al., 2020)

CHAPTER THREE

RAPID AUTHENTICATION OF COFFEE BEAN VARIETIES OF DIFFERENT FORMS BY USING POCKET-SIZED NIR SPECTROSCOPY AND MULTIVARIATE DATA MODELLING

*Vida Gyimah Boadu[1,2], *Ernest Teye[1], Charles L. Y. Amuah[3] & L. K. Sam-Amoah[1]

[1]University of Cape Coast, College of Agriculture and Natural Sciences, School of Agriculture, Department of Agricultural Engineering, Cape Coast, Ghana

[2]Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development, Ghana

Department of Hospitality and Tourism Education, Kumasi, Ghana

[3]University of Cape Coast, College of Agriculture and Natural Sciences, School of Physical Sciences, Department of Physics, Cape Coast, Ghana

*Corresponding authors

Published with slight modification, Journal of Analytical Methods (Scopus/ Sci, IF = 3.1)

Statement of contribution of Joint Authorship

Boadu, V. G.: **(Candidate)**

Conceptualised the topic, established methodology, data collection and analysis, preparation of tables and figures, writing and compilation of the original manuscript.

Teye, E.: **(Principal Supervisor)**

Conceptualised the topic, established methodology, supervised and edited the manuscript and co-author of manuscript.

Sam-Amoah, L. K.: **(Co-Supervisor)**

Conceptualised the topic, supervised and edited the manuscript and co-author of manuscript.

Charles L. Y Amuah: **(Research Team member)**

Validation, visualization and data curation and co-author of manuscript

Abstract

Coffee is the most consumed beverage and the second valuable traded commodity in the world. In this current study, a pocket sized near-infrared (NIR) spectroscopy and multivariate analyses were used for rapid authentication of coffee varieties (Arabica and Robusta) in three states to check mislabelling (Food Fraud). Two main coffee varieties were collected from different locations in Africa. The samples were scanned in 740 – 1070 nm wavelength and the spectral data were pre-treated with mean centring (MC), multiplicative scatter correction (MSC) and standard normal variate (SNV) independently while partial least square discriminate analysis (PLS-DA), K- nearest neighbour (KNN) and support vector machine (SVM) were used to comparatively build the prediction models for coffee beans (raw, roasted and powered). The performances of the models were evaluated by accuracy and efficiency. Among the classification methods developed, the best results were obtained for the following; raw coffee beans SD-SVM had accuracy of 0.92 and

efficiency of 0.82. For roasted coffee beans, SD-KNN had accuracy of 0.92 and efficiency of 0.87 while for roasted powdered coffee; FD-KNN showed an accuracy of 0.97 and efficiency 0.97. These finding reveals that for a more accurate differentiation of coffee beans, the roasted powder offers the best results. The obtained results show that pocket-sized NIR spectroscopy coupled with chemometrics could be employed to provide accurate and rapid authentication of different categories of coffee bean varieties.

Keywords: Coffee, Near infrared Spectroscopy, Arabica, Robusta, Multivariate algorithms.

3.1 Introduction

Coffee is one of the world's most important raw materials within international trade and the most popular beverage. Globally, coffee consumption for coffee year 2020/21 was estimated at 167.26 million bags, an increase of 1.9% as compared to last year's consumption (ICO, 2021). Brazil is the world's leading producer of coffee with 58 million coffee bags followed by Vietnam and Columbia. In Africa, the leading coffee producer is Ethiopia with 700 thousand of 60kg bags followed by Uganda and Ivory Coast (Zerga & Tsegaye, 2020). It is estimated that about 40 million people in Africa derive their livelihood from coffee. Among the numerous species of the genus coffee - coffee Arabica (Arabica coffee) and coffee canephora (Robusta coffee) are the two varieties of economical and commercial importance. Arabica coffee is the most prevalent and extensively cultivated species in the world, dominating 70% of the whole coffee production and 90% of the world

market (Zerga & Tsegaye, 2020). The two varieties differ not only in their botanical, chemical and sensory characteristics but also in commercial value, with Arabica coffee achieving higher market prices of 20-25% than Robusta (Buratti et al., 2015). Currently, the selling of coffee is based on its varietal and/or geographical origin (Martın et al., 1999). The international coffee trade is conducted in the raw state where Arabica and Robusta are easily distinguished by their appearance (e.g., size, shape and colour) by trained personnel. This visual criterion is destroyed once roasted and / or ground, the form in which coffee is commercially accessible to consumers (Casal et al., 2000). Presently most commercially available coffee beverages are produced from Arabica and Robusta roasted beans or blends of the two varieties. This has increased the possibility of disguising the coffee variety or accidental mislabelling, a subject of concern to food regulatory authorities and food manufactures (Downey & Boussion, 1996).

Different coffee samples differ in efficient antioxidants such as chlorogenic acids, phenolic acids, polyphenols and alkaloids and they are decomposed when roasted at temperatures (Dybkowska et al., 2017). However, the antioxidant content and efficiency can be preserved or improved by the development of compounds with antioxidant activity such as Maillard reaction products (Hečimović et al., 2011).

Over the few years, analytical methods used for authentication of coffee variety: solid phase microextraction−gas chromatography (SPME−GC) (Bicchi et al., 1997), HPLC analysis ((Farah et al., 2006), wet analytical method (Franca et al., 2005), reversed phase and normal phase high resolution liquid chromatography

(González et al., 2001) electrospray ionization mass spectrometry (ESI-MS) (Mendonça et al., 2009; Mendonça et al., 2008) liquid chromatography–mass spectrometry (Perrone et al., 2008), although consistent and precise, they are often expensive, time consuming, tedious, destructive and involves complex sample preparation. It is therefore very vital to develop a fast, efficient and onsite technique to overcome these short comings. Near Infrared spectroscopy has proven to be one of the most efficient and advanced tools to provide the solution.

Near-infrared (NIR) spectroscopy is increasingly becoming an extensive technique owing to its rapidity, simplicity, safety and non-invasive evaluation of food. It is currently identified to be an advanced and excellent technique for both qualitative and quantitative analysis in food and other industries. It has been extensively used in a range of fields, including agriculture, food chemistry, medicine, environmental science, petrochemistry and pharmaceuticals. It employs energy of wavelength in the range of 750 – 2500nm. Being a measurement technique, its advantages include rapid detection, potability, low-cost analysis and non-destructive measurement procedure with little or no sample preparation. However, the main drawbacks are the non-selectivity, the need for a precise and vigorous calibration with large data sets and the difficulty in transferring calibration between instruments (Páscoa et al., 2015). However, it offers the best option for onsite, rapid and user-friendly detection techniques.

This study seeks to authenticate coffee varieties from different African countries using pocket size spectroscopy. Several studies have used NIR spectroscopy to investigate the authentication of coffee and coffee products. Santos

et al. (2012) used NIR for classifying coffee samples from different lots and producers acquired in supermarket and roasting industries with wavelengths of 800 to 2500nm. Esteban-Díez et al. (2007) also used NIR spectroscopy to discriminate between Arabica and Robusta and a blend of the varied varietal composition between a range of 1100 to 2500nm. Buratti et al. (2015) used NIR spectroscopy to classify ground green coffee beans, ground roasted coffee and coffee beverages for both Arabica and Robusta coffee). Other use NIR to discriminate non defective and defective beans (Craig et al., 2012), assess other quality and adulteration parameters (Ebrahimi-Najafabadi et al., 2012); (Morgano et al., 2008; Pizarro et al., 2004); (Alessandrini et al., 2008), and predict the effect of roasting conditions (Lyman et al., 2003).

Food authentication is a key issue that has become increasingly significant in recent years. Certainly, quality is a major aspect for modern coffee industry because high quality product is the basis for success in today's competitive market. Consumers are interested in a reliably high-quality product at a reasonable price. It is therefore very important to develop a rapid, efficient, onsite technique for discriminating the two major varieties of coffee at the producing points and industries. To the best of our knowledge, up until now pocket size NIR spectroscopy has not been used to discriminate Arabica and Robusta in three different states (green coffee bean, roasted coffee bean and powered coffee bean). This study therefore seeks to evaluate a pocket-sized near infrared spectroscopy device for discriminating Arabica coffee and Robusta coffee at different states namely; raw, roasted and powdered. In this research, pocket-sized NIR

spectroscopy and multivariant data analysis were examined as a rapid and non-destructive method to authenticate different varieties of raw, roasted and powdered coffee.

3.2 Methodology

3.2.1 Coffee sample preparation

Two varieties of coffee bean samples (total 260), made up of 130 Robusta coffee and 130 Arabica samples each were collected from different locations in African coffee producing countries. They were collected from August to November, 2020-2021 season. The coffee bean samples used for this study were well dried, free from abnormal foreign odours and living insects. There was no evidence of adulteration, broken beans, pieces of shell foreign materials and were uniform in size. The well prepared and labelled samples were transported to the department of Agricultural Engineering laboratory of Cape Coast for examination.

3.2.2 Coffee beans roasting and grinding

The individual coffee bean samples were roasted separately according to (Vasconcelos et al., 2007) at a temperature of 200^{o}C for 1 h. For comparison purposes, roasting of coffee beans was conducted at the same temperature and power. Roasting of coffee was conducted with a home electric coffee bean roaster (JIAWANSHUN, China) which can contain 750 grams of coffee at a time. The same amount of roasted coffee beans of the two varieties were ground separately by a multi-purpose grinder (QE-100, Zhejiang YiLi Tool Co, Ltd, China) in the form of powder. All samples were kept in zip-locked bags for further analysis.

3.2.3 Reference measurement of coffee bean quality

Some chemical properties such as lipids, moisture content, ash, proteins, carbohydrates, fibre, polyphenols and antioxidants of coffee beans from the various samples collected were measured (triplicate) by using standard recommendation references analytical methods (AOAC, 1990).

3.3 Spectra Acquisition

A pocket-sized NIR spectroscopy (SCIO^{TM} UK) was used to scan the raw beans, roasted beans and powered coffee samples. The spectral data were acquired at 1nm resolution over a wavelength range of 740 – 1070 nm. The pocket-sized spectroscopy gives the measurement in relative absorbance unit (log 1/R), which were associated with chemical constituents. The device was operated using a smart phone (Samsung Galaxy A21) application with spectral data stored remotely. All coffee bean samples were carefully scanned five times through a transparent zip locked polyethene bag at different positions. The five scans were averaged for a sample. The transparent bag showed no significant interferences with the NIR signals

3.4 Spectral Data Partition

The spectral data-set for all the green beans, roasted beans and powered coffee (total 780 samples) were downloaded separately and each category was partitioned into two subsets: training set and prediction set. For all the coffee bean samples, 175 samples and 85 samples were selected as the training set and testing set respectively. The training set was used to develop the model while the testing

set was used to evaluate the actual predictability of the model. The individual samples in each set were selected randomly in order to come to a 2/1 division of the training set and testing set as was done by other authors (Teye et al., 2014).

3.5 Theories of Spectra Pre-Processing

All the computations, chemometric analysis and graphics were done with MATLAB (The MathWorks, Inc., USA, version 9.6.0.) using the windows 10 Pro software package. In this studies, different pre-processing method was done to ascertain its influence and improve the raw dataset. The pre-processing methods employed were: first derivative (FD), second derivative (SD), mean centring (MC), multiplicative scatter correction (MSC) and standard normal variant (SNV). After a trial-and-error procedure, all the five methods were selected.

First derivative (FD) pre-processing is very effective for removing baseline offset. Spectral derivative transformation is one of the best methods for removing baseline defects (Duckworth, 2004).

Second derivative (SD) pre-processing technique removes baseline drifts, the linear trend from a spectrum (Duckworth, 2004; Rinnan et al., 2009) and detached overlapping peaks while small spectral differences are improved**.** The application of derivative chemometric tools helps to detect small chemical changes in the samples

Mean centring (MC) pre-treatment method is achieved by subtracting the average values from the spectrum from the calculated average spectrum of the data set. This

was done to ensure interpretable results as performed by Teye and co-workers who deployed this as pre-treatment technique (Teye et al., 2013).

Multiplicative scattered correction (MSC) is a special pre-treatment method used for the correction of scattered light on different particle size. The working principle is to correct additive and multiplicative effects in the spectra. It is also useful for eliminating or minimizing un-useful variability owing to scattering, that is, it decreases the effects of light emission and reduces diversity in the spectrum. MSC assumes that all samples have the same scatter level as a reference spectrum. The procedure is to study each spectrum and correct it based on the reference spectrum by utilizing the results of simple linear regression estimation (Dhanoa et al., 1994).

Standard normal variant (SNV) transformation is applied to individual spectrum by subtracting the spectrum mean and dividing by the spectrum standard deviation (Candolfi et al., 1999). SNV corrects the multiplicative interferences of light scatter and particle size and the change light distance. Each object was transformed independently to remove slope variation on the individual spectrum basis. Both additive and multiplicative scatter effects are corrected by SNV (Dhanoa et al., 1994).

3.6 Principal Component Analysis (PCA)

This was conducted after pre-processing the spectral profile. PCA is an unsupervised pattern regulation method which is used for visualizing data trends in dimensional space. It is a versatile method capable for revealing relations between variable, cluster samples, detecting outliners, finding and quantifying patterns as

well as generating hypothesis (Bro & Smilde, 2014). It works by reducing the dimensions of the data matrix and compresses the information into few interpretable variables called principal components (PCs), which are linear combinations of the original variables (Luna et al., 2013). The best PCs usually show the most vital information. Hence, similar samples are grouped closer to each other and vice versa. The graphical outline of PCA results usually gives preliminary output for the determination of possible variance and similarities in the data set. PCA can be used to detect combinations of variations in the dataset, as these variables are kept in the first two or three PCs and the PCA loading plot explain these contributions.

3.7 Multivariate Classification Algorithms

After applying the above-mentioned spectral pre-processing techniques, different multivariate classification algorithms were studied systematically and their results were compared. These classification methods include partial least square discriminate analysis (PLSDA), K-nearest neighbour (KNN) and support vector machine (SVM). For more information on the theories of the applied multivariate algorithm refer other authors (Cortes & Vapnik, 1995; Khamar, 2013; Lee & Jemain, 2019; Lee et al., 2018; Tang et al., 2014).

3.8 Results and Discussion

3.8.1 Wet chemistry results

The average data of the biochemical composition of the coffee samples (green and roasted) from the two main varieties; Robusta and Arabica are shown in

Table 3.3. Reported values of proximate composition for raw and roasted coffee of the two varieties are in the same range as reported by literature (Franca & Oliveira, 2008; Gyedu-Akoto et al., 2019a; Vasconcelos et al., 2007).

Moisture content for the two varieties of the raw and roasted coffee are in the range from 7 to 8% and 4 to 5% respectively. The recorded moisture content is favourable, as it will reduce the growth of mycotoxin and prevent changes in physical, chemical and sensory parameters. The ash content of the raw coffee was in the range 4 to 5% and was reduced to 3 to 4% when roasted which is an indication of the mineral content, quality and safety. Protein, lipids, carbohydrates of raw coffee were in the ranges of 14 to 18%, 3 to 7% and 67 to 70% respectively. There was slight decrease in protein (12 to 16%), carbohydrates (65 to 66%) and slight increase in lipids (6-8%) of the roasted coffee. All the values for coffee (green and roasted) for the two varieties were in range as reported in literature (Franca & Oliveira, 2008; Gyedu-Akoto et al., 2019a; Vasconcelos et al., 2007).

There was a decrease in the concentration of polyphenols from 2349 to 1019 mg kg^{-1} for Arabica and 2415 to 1039 mg kg^{-1} for Robusta coffee. Polyphenols are degraded in the application of heat to food products. However, there was a significant rise in antioxidant activity from 3200 to 3286 mg kg^{-1} for Arabica and 3675 to 3739 mg kg^{-1} for Robusta coffee. This could be as a result of the formation Maillard reaction compounds during roasting where antioxidants melanoidin compounds are formed which compensate for the decrease in polyphenol (Dybkowska et al., 2017). Although natural antioxidants are lost, antioxidant activities of roasted coffee increased. The differences observed for the two unique

varieties are cumbersome if you depend only on the proximate composition. Therefore, this research further employed non-destructive techniques as in the case of NIR spectroscopy.

Table 3.3: Biochemical parameter examination results for green and roasted coffee beans

Parameter	Variety	Green Arabica beans	Green Robusta beans	Roasted Arabica beans	Roasted Robusta beans
Moisture (%)	Average	7.700	8.991	5.323	4.752
	SDV	0.186	0.406	0.217	0.198
Ash (%)	Average	4.460	5.100	3.897	4.414
	SDV	0.257	0.218	0.106	0.067
Protein (%)	Average	14.763	18.606	12.265	16.243
	SDV	0.144	0.069	0.232	0.336
Lipid (%)	Average	7.243	3.732	8.442	6.754
	SDV	0.100	0.128	0.060	0.268
CHO (%)	Average	70.039	67.320	66.248	65.029
	SDV	0.125	0.066	0.320	0.321
Fibre (%)	Average	7.020	6.456	5.474	5.356
	SDV	0.022	0.139	0.231	0.110
Polyphenols (mg kg^{-1})	Average	2349.83	2415.60	1019.51	1039.65
	SDV	0.007	0.008	0.020	0.012
Antioxidants (mg kg^{-1})	Average	3200.23	3675.42	3286.72	3739.39
	SDV	0.037	0.017	0.042	0.152

3.8.2 Spectral profile

It could be observed (Figure 3.4) that the original spectra of the green coffee bean changes with different pre-processing techniques and each showed a unique profile. Figure 3.4 presents the spectral profile of the selected pre-processing techniques applied on green coffee beans for Arabica and Robusta from different countries. From the spectral profile, it was detected that different pre-processing methods exhibited their advantage for different investigated challenges to give the optimum results. The absorbance data for this novel investigation was obtained

with a wavelength of 740 – 1070 nm. This wavelength range can offer a significant feature for classification of different coffee bean variety. It is well known that Arabica coffee has different composition from Robusta coffee bean in terms of amount of carbohydrates, proteins, lipids, moisture, polyphenols, antioxidants, volatiles and non-volatiles compounds as recorded for the biochemical parameters. When both are radiated to light with various wavelengths, there is absorption of a portion of the light at certain wavelength. The amount of light absorbed for each variety depends on the aforementioned composition of each coffee beans which shows a number of bands and peaks on the spectral profile. Spectroscopic analysis is characterised by a larger number of variables, both useful and non-useful information. Multivariate calibration techniques were then used to extract this useful information.

The spectra of roasted coffee (Arabica and Robusta) shares similar absorbance values of spectra. This could be explained as a result of roasting the coffee bean to become dry, brown, brittle, increase in size, develop flavour and aroma when roasted at 200ºC. Water and carbon dioxide are released while volatile aromatic compounds are formed (Oestreich-Janzen, 2013). After roasting of the coffee beans, the amount of carbohydrates, chlorogenic acids and amino acids are reduced in quantity while the amount of antioxidants are improved as a Milliard reaction products (Hečimović et al., 2011).

It was observed in Figure 3.5 (a) that the space between the spectral profile for Arabica and Robusta coffee widens from wavelength 900nm to 1080nm for the roasted coffee beans. This could be due to the higher concentration of

carbohydrates, lipids and trigonelline in roasted Arabica compared to the other as well as the higher levels of caffeine in roasted Robusta coffee (Souza & Benassi, 2012).

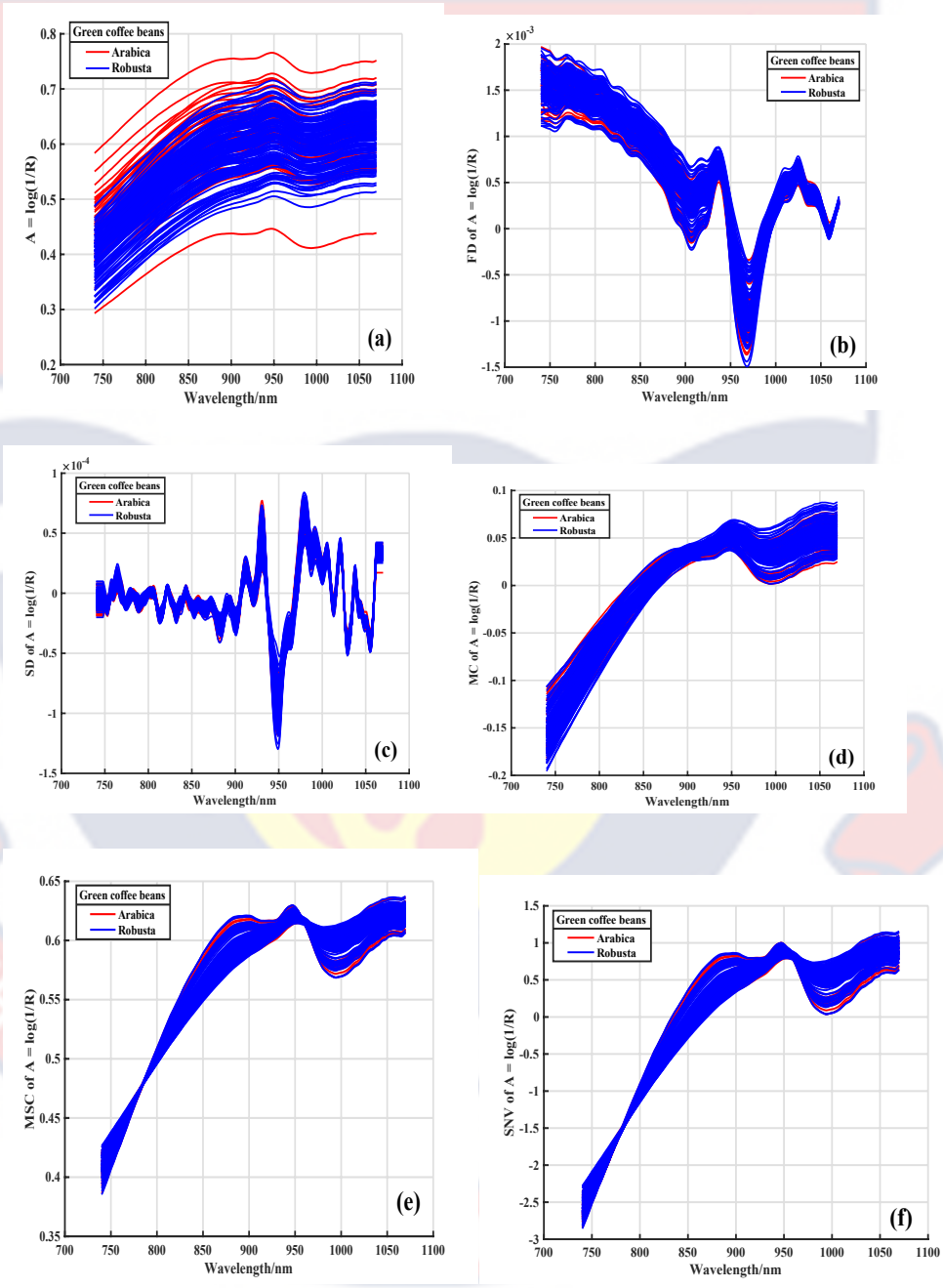


Figure 3.4: Spectral profile of green coffee beans (a) raw, (b) FD, (c) SD (d) MC (e) MSC and (f) SNV

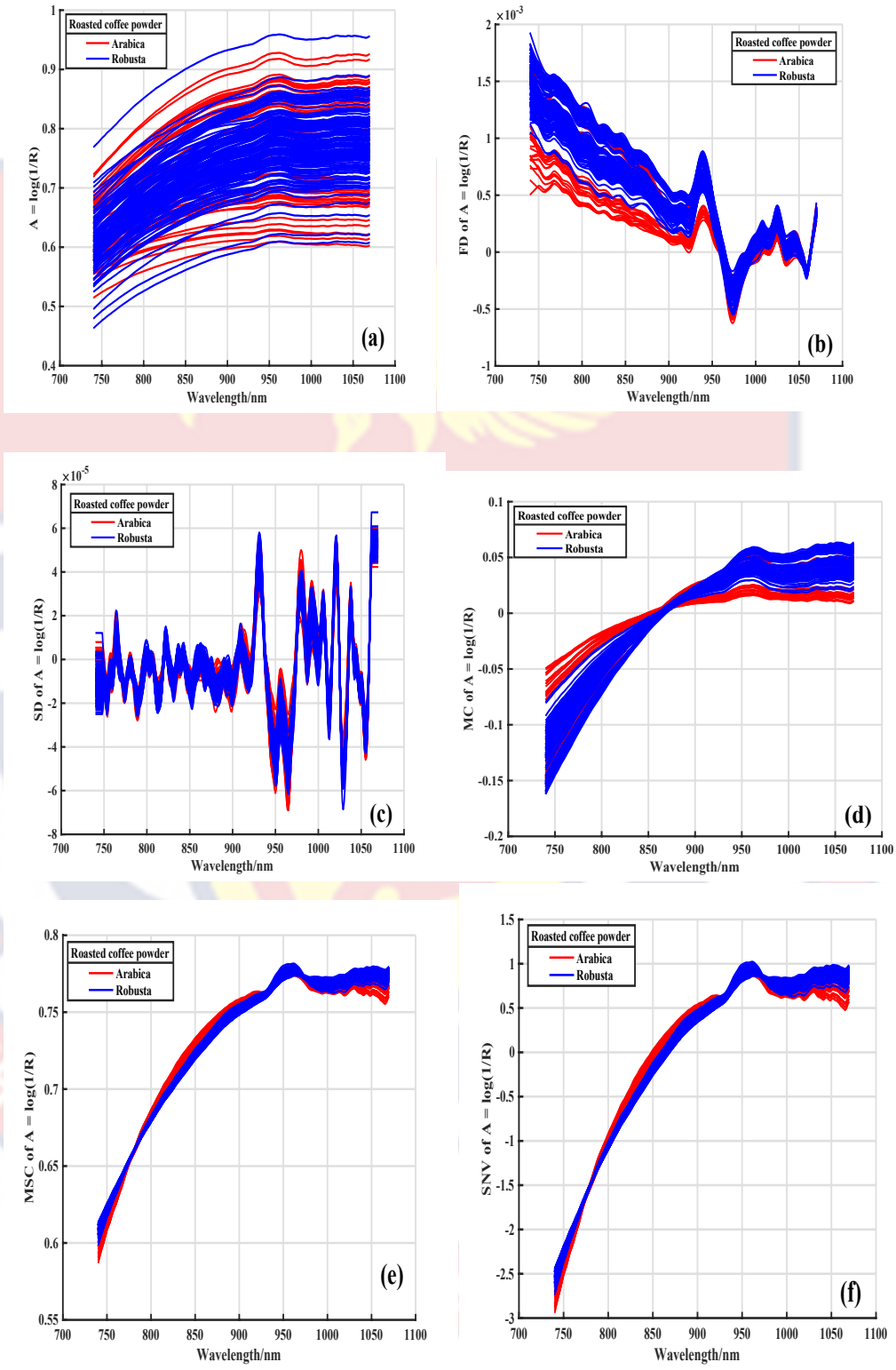


Figure 3.5: Spectral profile of roasted coffee powder (a) raw, (b) FD, (c) SD, (d) MC, (e) MSC and (f) SNV

3.8.3 Principal component analysis (PCA)

3.8.3.1 Raw coffee bean

PCA is an unsupervised pattern recognition method which is used for visualizing data trends in a dimensional space. Different pre-processing techniques were used for classifying two varieties (Arabica and Robusta) of raw coffee samples with their PCA from different countries in Africa. PCA can provide a spectral information trend but it is not a classification tool. The topmost three principal component (PC1, PC2 and PC3) obtained from a scatter plot were used to ascertain the clear cluster trend of a sample as seen from Figure 3.6. For MC the total contribution of PCs was 99.8% (PC1 = 91.77, PC2 = 7.72, PC3 =0.02), for MSC PCs contributed to 99.63% (PC1 = 97.69, PC2 = 1.54, PC3 = 0.40) and PCs for SNV contributed to 99.35% (PC1 = 97.15, PC2 = 1.79, PC3 = 0.41) of the total variance in the dataset of the coffee varieties. This means that all the three preprocessing techniques had PCs that can explain 99% of the variance data from the spectral information which covers the relevant chemical fingerprints of the coffee bean sample used.

The PCA results discovered that MC provides the best pretreatment results compared with the others because it had the highest value and a well-ordered grouping. The differences in the chemical, physical and biochemical parameters of the two varieties have resulted in the grouping of the raw coffee beans. For example, Arabica coffee are: larger in size with high amount of carbohydrate, amino acids, organic acids and lipids while Robusta are: smaller in size with high amount of caffeine, chlorogenic acids and ash (Oestreich-Janzen, 2013).

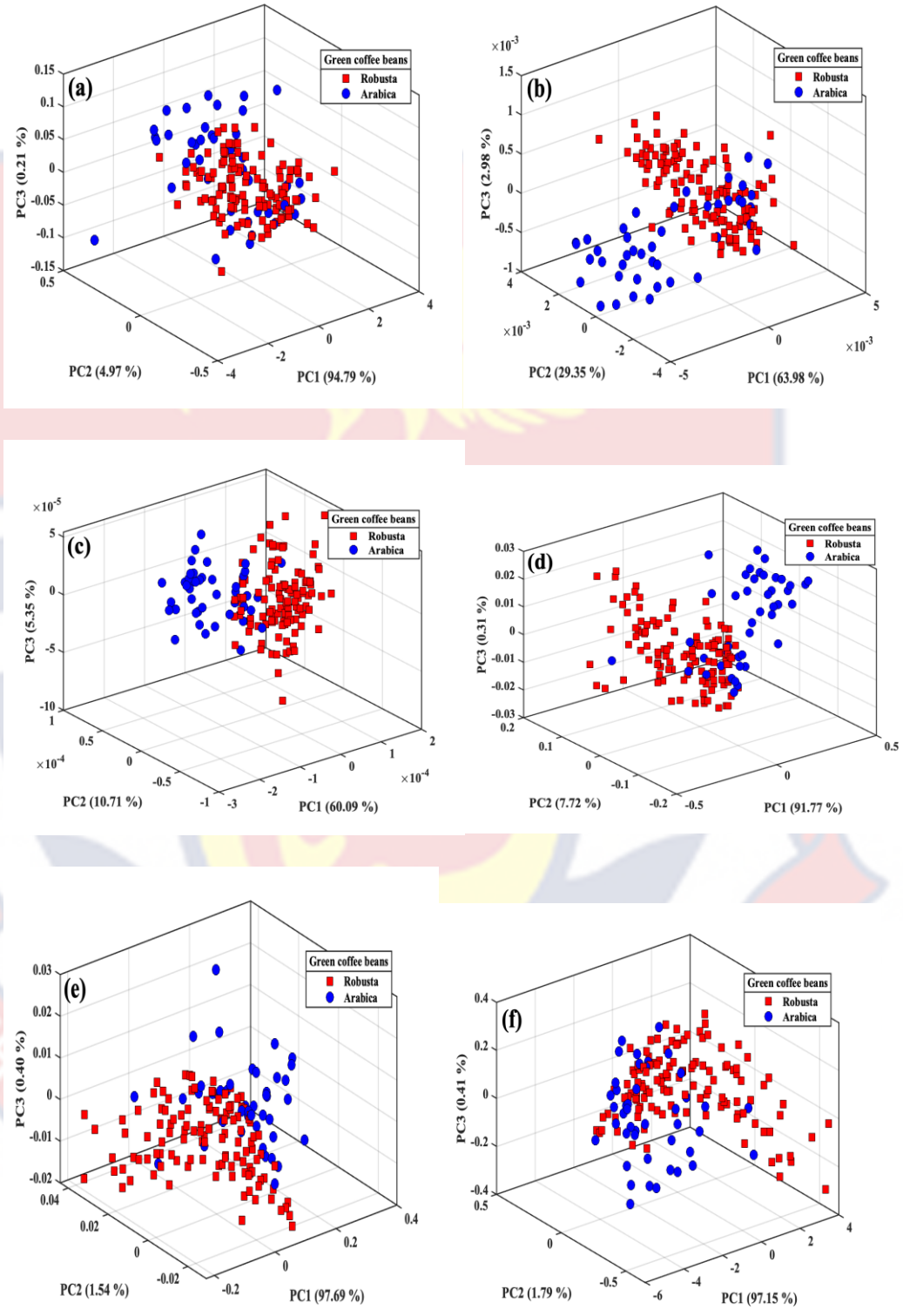


Figure 3.6: PCA score plot of the first three PCs of green coffee bean for Arabica and Robusta (a) raw, (b)FD, (c) SD, (d) MC, (e) MSC, (f) SNV

3.8.3.2 Roasted coffee bean

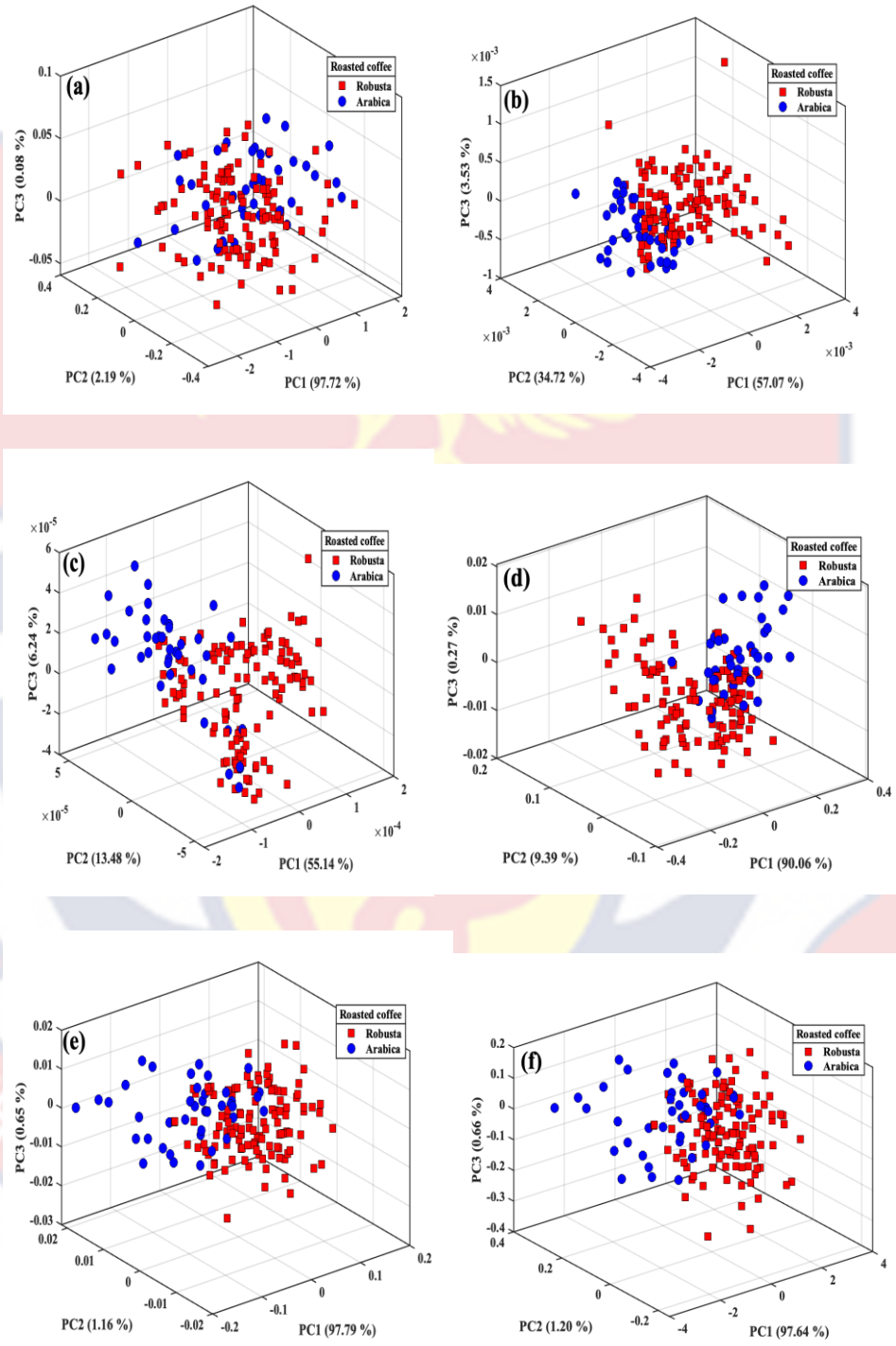


Figure 3.7: PCA score plot of the first three PCs of (a) raw, (b) FD, (c) SD, (d) MC, (e) MSC, (f) SNV roasted coffee for Arabica and Robusta

For roasted coffee beans, PCs of MC contributed to 99.72% MSC with 99.6% and SNV with 99.5% of the total variance in the dataset of the roasted powdered coffee for the two varieties as shown in Figure 3.7. This means that these three pre-processing techniques can explain 99.72% of the variance information from the spectral data that covers the vital chemical fingerprints in the coffee samples used. The PCA results discovered that MC provides the best pre-treatment results of 99% compared with the others.

3.8.3.3 Roasted coffee powder

Also, for roasted coffee bean and powdered, it could be seen from Figure 3.8 that, MC PCs contributed to 99.92%, MSC PCs with 98.72% and SNV PCs with 98.39% of the total variance in the dataset of the roasted powdered coffee for the two varieties. This means that these, MC, MSC and SNV can explain 98% of the variance information from the spectral data that covers the vital chemical fingerprints in the powdered coffee samples used. The PCA results discovered that MC provides the best pretreatment results of 99% compared with the others.

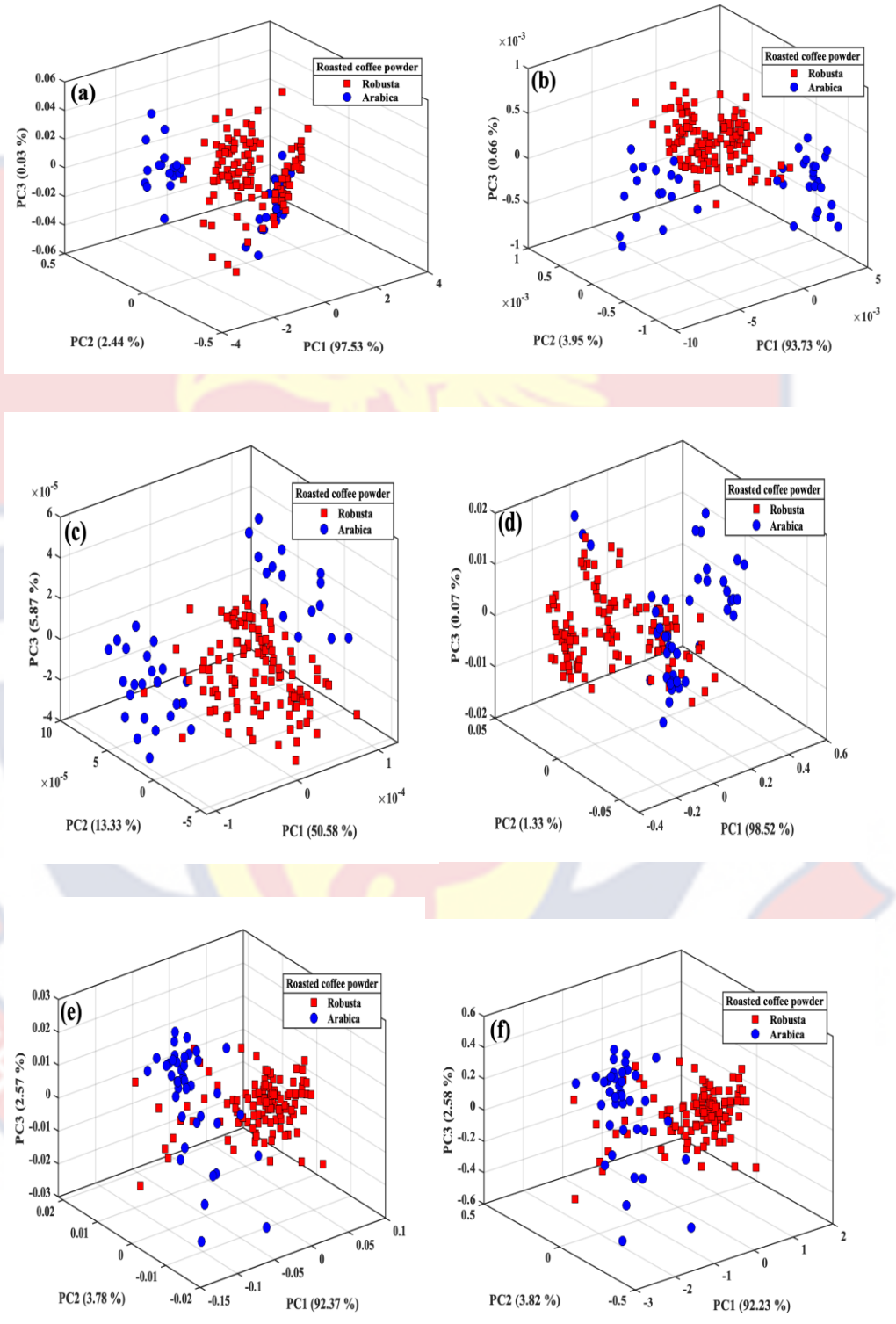


Figure 3.8: PCA score plot of the first three PCs of roasted coffee powder for Arabica and Robusta (a) raw, (b) FD (c) SD, (d) MC, (e) MSC, (f) SNV

3.8.4 Mean spectral and PCA loading plot

To appreciate the separation of the different coffee samples, mean spectral profile and PCA loading plot was done as shown in Figure 3.9. Concerning the raw mean spectra, there was an overlapping of both spectra from 950 nm to 980 nm of the roasted powdered coffee. This could be due to loss of volatile aromatic compounds such as furan and optimization of antioxidant properties in both coffee variety. It could be detected that the significant absorption bands for the first component were around 950 nm, 1000 nm and 1100 nm as seen in Figure 3.9. These bands represent -CH_3, CH_2 3rd overtone, RNH, ROH and H_2O 2nd overtone respectively which could correspond to monosaccharides, oligosaccharides, polyphenols and antioxidants properties (Barbin et al., 2014).

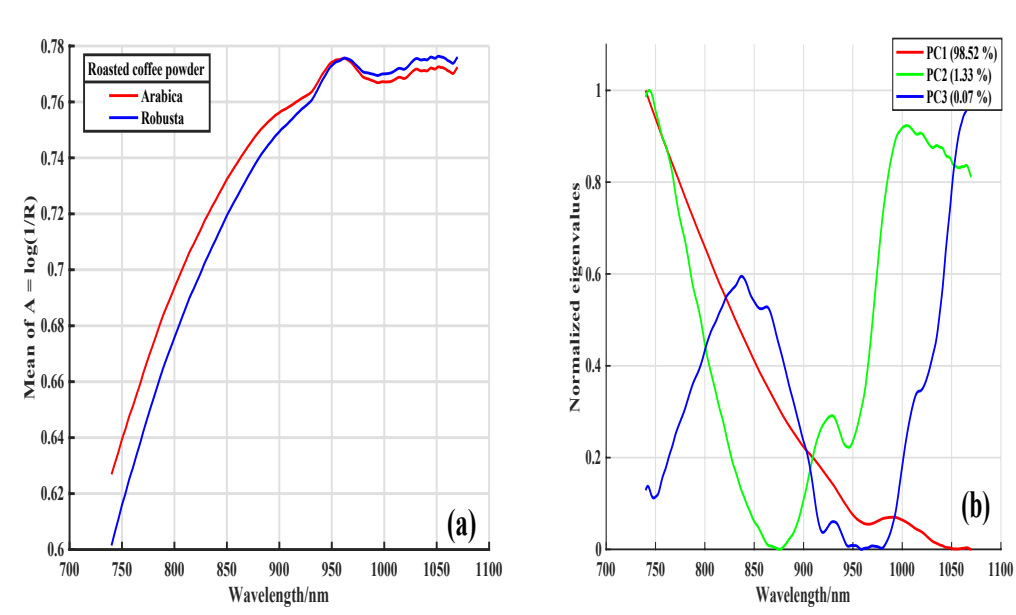


Figure 3.9: Spectral profile (a) raw mean and (b) MC loading plot for roasted coffee powder

3.8.5 Classification based on states of coffee bean variety

3.8.5.1 Raw coffee beans

For the rapid classification of the two varieties, PLSDA, KNN and SVM were separately applied after PCA to authenticate them at different forms (raw, roasted and powder) of the coffee bean variety. It was observed in Table 3.4 that, SD-SVM model gave the best classification rate with accuracy of 0.92 and efficiency of 0.82 for raw coffee bean. The model was also considered as the best for this category, due to the high values for specificity and precision. The values obtained for moisture, ash, proteins, lipids, carbohydrates, polyphenols and antioxidants for the raw coffee bean were in accordance with data obtained by other authors (Oestreich-Janzen, 2013; Oliveira et al., 2006). Although there was no significant difference among the parameters measured, the pocket-size spectroscopy was able to differentiate the green coffee bean varieties based on the analysed spectroscopic data. The quality of the biochemical parameters' measurement contributed to the differentiation of the spectra data.

Table 3.4: Performance of classification algorithms with different pre-processing techniques of green coffee beans

Model		Predict rate	Accuracy	Error	Sensitivity	Specificity	Efficiency	Precision
PLSDA	Raw	87.73	0.87	0.12	0.53	0.98	0.72	0.90
	MC	90.79	0.90	0.09	0.71	0.97	0.83	0.89
	MSC	89.57	0.89	0.10	0.58	1.00	0.76	1.00
	SNV	88.95	0.88	0.11	0.54	1.00	0.74	1.00
	FD	90.18	0.90	0.09	0.61	1.00	0.78	1.00
	SD	87.11	0.89	0.10	0.79	0.93	0.86	0.79
KNN	Raw	87.11	0.85	0.14	0.47	1.00	0.68	1.00
	MC	89.57	0.92	0.08	0.67	1.00	0.81	1.00
	MSC	87.73	0.90	0.09	0.63	0.99	0.79	0.95
	SNV	87.73	0.90	0.09	0.63	0.98	0.79	0.95
	FD	88.95	0.90	0.09	0.67	0.99	0.81	0.96
	SD	89.57	0.92	0.07	0.87	0.97	0.87	0.89
SVM	Raw	86.50	0.87	0.12	0.54	0.98	0.73	0.90
	MC	89.57	0.89	0.10	0.57	1.00	0.75	1.00
	MSC	89.57	0.89	0.10	0.57	1.00	0.76	1.00
	SNV	89.57	0.90	0.09	0.61	1.00	0.79	1.00
	FD	89.57	0.89	0.10	0.57	1.00	0.76	1.00
	***SD**	**91.41**	**0.92**	**0.08**	**0.67**	**1.00**	**0.82**	**1.00**

***Best model**

3.8.5.2 Roasted coffee bean

After the classification of green coffee beans for the two varieties, the possibility of distinguishing the roasted forms of coffee beans for Arabica and Robusta was studied. By analysing the results as seen in Table 3.5, it was observed that SD-KNN model performed better for the classification of roasted coffee beans with an accuracy of 0.92 and efficiency of 0.87.

Table 3.5: Performance of classification algorithms (with different pre-processing techniques of roasted coffee beans

Model		Predict rate	Accuracy	Error	Sensitivity	Specificity	Efficiency	Precision
PLSDA	Raw	78.52	0.77	0.22	0.34	0.93	0.57	0.67
	MC	79.75	0.78	0.21	0.37	0.93	0.59	0.68
	MSC	82.20	0.83	0.16	0.54	0.93	0.71	0.71
	SNV	81.59	0.83	0.16	0.54	0.93	0.71	0.71
	FD	82.82	0.86	0.13	0.54	0.96	0.73	0.81
	SD	88.95	0.90	0.09	0.74	0.95	0.84	0.87
KNN	Raw	77.91	0.77	0.22	0.43	0.87	0.62	0.54
	MC	80.36	0.80	0.19	0.44	0.93	0.64	0.71
	MSC	84.43	0.83	0.16	0.64	0.90	0.77	0.71
	SNV	84.43	0.83	0.16	0.59	0.90	0.73	0.68
	FD	76.68	0.77	0.22	0.13	0.95	0.36	0.44
	***SD**	**88.95**	**0.92**	**0.07**	**0.77**	**0.97**	**0.87**	**0.89**
SVM	Raw	78.52	0.75	0.24	0.22	0.95	0.46	0.67
	MC	81.59	0.81	0.18	0.51	0.90	0.68	0.60
	MSC	82.20	0.81	0.18	0.58	0.90	0.73	0.70
	SNV	84.04	0.83	0.16	0.51	0.93	0.69	0.68
	FD	80.36	0.77	0.22	0.33	0.93	0.56	0.67
	SD	86.50	0.87	0.12	0.55	0.97	0.73	0.84

***Best model**

3.8.5.3 Roasted coffee powder

The performance of the classification for roasted coffee powder is summarised in Table 3.6. FD-KNN performed comparatively better than the others with accuracy of 0.97 and efficiency of 0.97 for roasted powdered coffee. This model was also considered to the best among the other models for the classification of the powdered coffee due to high sensitivity, specificity, efficiency and precision. This means that pocket size spectroscopy could be useful to differentiate coffee in different forms; green, roasted and roasted coffee powder for both Arabica and Robusta coffee. Also, it can be suitable to differentiate the vital quality attribute of the varieties of coffee in addition to the observed chemical properties.

Table 3.6: Performance of classification algorithms with different pre-processing techniques of roasted coffee powder

Model		Prediction rate	Accuracy	Error	Sensitivity	Specificity	Efficiency	Precision
PLSDA	Raw	89.57	0.90	0.09	0.77	0.94	0.85	0.80
	MC	86.50	0.87	0.12	0.80	0.89	0.85	0.69
	MSC	87.73	0.87	0.12	0.56	0.97	0.74	0.90
	SNV	92.63	0.95	0.04	0.83	0.99	0.91	0.96
	FD	85.88	0.87	0.12	0.83	0.88	0.86	0.68
	SD	81.59	0.82	0.17	0.64	0.88	0.75	0.63
KNN	Raw	96.93	0.96	0.03	1.00	0.96	0.98	0.89
	MC	95.09	0.95	0.04	0.93	0.96	0.95	0.88
	MSC	95.70	0.96	0.03	0.93	0.98	0.96	0.94
	SNV	95.70	0.96	0.03	0.93	0.98	0.96	0.94
	***FD**	**96.93**	**0.97**	**0.02**	**0.96**	**0.98**	**0.97**	**0.94**
	SD	95.70	0.96	0.03	0.93	0.97	0.95	0.90
SVM	Raw	90.18	0.90	0.09	0.75	0.95	0.85	0.87
	MC	90.79	0.90	0.09	0.80	0.94	0.87	0.85
	MSC	93.25	0.93	0.06	0.86	0.95	0.91	0.89
	SNV	93.25	0.93	0.06	0.86	0.95	0.91	0.89
	FD	85.88	0.87	0.12	0.77	0.91	0.84	0.77
	SD	84.04	0.82	0.17	0.47	0.95	0.67	0.80

***Best model**

3.8.6 General discussion

It was observed that the original spectra of green coffee beans show a unique profile for Arabica and Robusta coffee samples. From the spectral profile, it could be observed that Arabica coffee has a different composition from Robusta coffee bean in terms of the amount of carbohydrates, proteins, lipids, moisture, polyphenols, antioxidants, volatiles and non-volatiles compounds. After roasting of the coffee beans, the amount of carbohydrates, chlorogenic acids and amino acids are reduced in quantity while the amount of antioxidants are improved as a Milliard reaction products (Hečimović et al., 2011). Water and carbon dioxide are released while volatile aromatic compounds are formed (Oestreich-Janzen, 2013). The raw mean spectra for Arabica and Robusta coffee widens for the roasted beans coffee.

This could be a higher concentration of carbohydrates, lipids and trigonelline in roasted Arabica and higher levels of caffeine in roasted Robusta coffee (Souza & Benassi, 2012). There was an overlap of spectra for roasted powder for Arabica and Robusta and this could be due to the loss of volatile aromatic compounds and optimization of antioxidant properties in both coffee varieties. MC was the best pre-treatment for the green, roasted and roasted coffee powder samples because it scores the highest values among the lot and also provided a well-ordered grouping by bringing the differences in the chemical, physical and biochemical properties of the two varieties.

For the classification model, PLSDA, KNN and SVM were separately applied after PCA to perform the classification among the different states/forms (green, roasted and powder) of coffee varieties. The best performance was recorded for the following in terms of accuracy and efficiency; SD-SVM had 0.92 and 0.82 for the green coffee beans, SD-KNN had 0.92 and 0.87 for roasted coffee bean and FD-KNN had 0.97 and 0.97 for roasted powdered coffee. These models were considered to be the best, however, SD-KNN was superior to all. This could be as a result of the release of more unique fingerprints during roasting and powdering as well as the strength of SD for detecting small changes in the spectra while KNN classifier helps to analysed the spectral differences. In the classification of green coffee, FD was able to remove baseline defect in spectra formation whiles SVM revealed good performance for the classification.

3.8.7 Conclusion

This work suggests that pocket-size near-infrared coupled with appropriate chemometric analysis is a rapid and non-destructive method for authenticating different categories of coffee variety. The methodical selection of different pre-processing techniques (FD, SD, MC, MSC and SNV) with PCA modelling together with PLSDA, KNN and SVM presented an advantage in authenticating the coffee varieties in three states (green, roasted and powder). The best performance was recorded for the following categories considering accuracy and efficiency; SD-SVM with 0.92 and 0.82 for the raw coffee beans, SD-KNN with 0.92 and 0.87 for roasted coffee bean and FD-KNN with 0.97 and 0.97 for roasted powdered coffee. Generally, it could be concluded that pocket size spectroscopy together with chemometrics could be used to exploit rapid authenticating of different varieties of coffee. Besides, these models can be imported into smartphones for effective and accurate authenticating of coffee. Further studies are needed to authenticate the regional and geographical origins of coffee beans.

CHAPTER FOUR

RAPID CLASSIFICATION OF AFRICAN GEOGRAPHICAL COFFEE TYPES BY HANDHELD NIR SPECTROSCOPIC METHOD

*Vida Gyimah Boadu[1,2], *Ernest Teye[1], Charles L. Y. Amuah[3] & L. K. Sam-Amoah[1]

[1]University of Cape Coast, College of Agriculture and Natural Sciences, School of Agriculture, Department of Agricultural Engineering, Cape Coast, Ghana

[2]Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development,

Department of Hospitality and Tourism Education, Kumasi, Ghana

[3]University of Cape Coast, College of Agriculture and Natural Sciences, School of Physical Sciences, Department of Physics, Cape Coast, Ghana

*Corresponding authors

Accepted for publication, Journal of Foods, (Scopus/Sci, IF = 5.5)

Statement of contribution of Joint Authorship

Boadu, V. G.: **(Candidate)**

Conceptualised the topic, established methodology, data collection and analysis, preparation of tables and figures, writing and compilation of the original manuscript.

Teye, E.: **(Principal Supervisor)**

Conceptualised the topic, established methodology, supervised and edited the manuscript and co-author of manuscript.

Sam-Amoah, L. K.: **(Co-Supervisor)**

Conceptualised the topic, supervised and edited the manuscript and co-author of manuscript.

Charles L. Y Amuah: **(Research Team member)**

Validation, visualization and data curation and co-author of manuscript

Francis Padi Lamptey: **(Research Team member)**

Validation, visualization and data curation and co-author of manuscript

Abstract

African coffee is among the best traded coffee type worldwide and rapid identification of their geographical origin is very important when trading the commodity. In this study, geographic differentiation of African coffee types (bean, roasted, and powder) was achieved using handheld near-infrared spectroscopy and multivariant data processing. Five African countries were used as the origins for the collection of Robusta coffee. After being individually preprocessed with mean centering (MC), multiplicative scatter correction (MSC), and standard normal variate (SNV), the samples were scanned at a wavelength of 740–1070 nm. Support vector machines (SVM), linear discriminant analysis (LDA), neural networks (NN), random forests (RF), and partial least square discriminate analysis (PLS-DA) were then used to develop a prediction model for African coffee types. The performance of the model was assessed using accuracy and F1-score. The best classification algorithms were developed for the following coffee types: raw bean coffee, SD-PLSDA, and MC+SD-PLSDA. These models had an accuracy of 0.87

and an F1-score of 0.88. SNV+SD-SVM and MSC+SD-NN both had an accuracy and F1 scores of 0.97 for roasted coffee beans and 0.96 for roasted coffee powder, respectively. The obtained results revealed that efficient quality assurance may be achieved by using handheld NIR spectroscopy in combination with chemometrics to differentiate between different African coffee types according to their geographical origins.

Keywords: Coffee bean, NIR spectroscopy, Partial Least Squares-Discriminant Analysis, Robusta, geographical differentiation, chemometrics

4.1 Introduction

Coffee is one of the most consumed beverages, widely traded commodity in the world, and a vital product for many developing countries. Globally, South America's coffee production is the highest at 42%, Africa at 20.4% and Asia at 18.5% (Martín et al., 2001). Coffee is grown in more than one hundred different tropical countries, engaging about 20million people in the production of 6.7million tons (ICO, 2004). Moreover, it is one of the important commodities that are traded worldwide, accounting for nearly half of all tropical product exports. (Hečimović et al., 2011). The only two cultivated and commercially available coffee beans among the 100 species in the genus of coffee are coffee arabica (Arabica) and coffee canephora (Robusta), accounting for 56% and 44% respectively. They are usually subjected to adulteration, mislabelling products to hide the true botanical and geographical origin due to large diffusion and high market value. Some coffees are most highly appreciated whereas others are considered of lower quality due to the methods of harvesting and processing methods used in the country (Agresti et al.,

2008). Arabica coffee has a more sweet and fruity flavour profile than Robusta coffee because of the differences in the environment (cultivated on slopes), growing conditions, and methods of processing and drying (Cagliani et al., 2013). The characteristics of the finished product are affected by technological processes such as roasting and grinding in addition to environmental factors and harvesting procedures. (Bressanello et al., 2017). Although Arabica coffee is generally regarded as superior to Robusta because of its fine and aromatic flavour, the latter in certain regions is a viable competitor due to lower production costs, higher yield, significantly lower price (Mendes et al., 2001) and higher amount of caffeine (DaMatta et al., 2007).

In Africa, coffee is mainly cultivated on small-scale farms with constrained and dispersed land holdings, insufficient access to inputs, and low prices. It is produced using a variety of farming techniques, mostly by interplanting trees for shade and other crops. However, coffee species are severely threatened by genetic erosion and irreversible loss in Africa's centres of origin and variety. This susceptibility is mostly brought on by factors such as population growth, the expansion of large farms, crop substitution, the coffee crisis, and climate change, among others (Kufa, 2010). According to Sylvain (1958), Robusta coffee comprises the majority of the coffee grown in Africa. Due to the rust disease and several insect pests in the lowlands, the situation for Arabica is not ideal, but some areas in the eastern Hemisphere are successful in their growth. Because Robusta blends well with other coffee varieties, the significant growth of

the soluble coffee industry in the United States has fortunately improved the market for this coffee variety.

Coffee plays an essential role in the national economies of these African countries; Ghana, Ivory Coast, Uganda, Kenya, Nigeria and others (ICO, 2009a). Robusta originates from the equatorial lowland forests of west and central Africa. Coffee producing countries with lowland areas such as Ghana, Nigeria, Ivory Coast and Burkina Faso produce Robusta coffee. Considering that pricing currently depends on geographic origin, determining the geographical origin of commodities and food products is becoming an increasingly active research area. (ICO, 2012). It has protected the market share, reputation, and customer trust to pay a premium for particular coffee-growing regions (Anderson & Smith, 2002). Furthermore, it has gained relevance because a variety of geographical areas have different biochemical and organoleptic characteristics. Due to the aforementioned socioeconomic concerns, coffee farmers and industrial manufacturers are becoming more concerned with maintaining the market's reputation and have strongly encouraged the development of effective analytical methods for determining the authenticity of coffee (Huck et al., 2005).

Different analytical methods are often employed for the differentiation of coffee samples of different origins such as gas and liquid Chromatography (Amalia et al., 2021; Núñez, Martínez, et al., 2021), wet chemistry (Zhu, Long, Ma, et al., 2021), mass spectroscopy (Gil-Agusti et al., 2005), UV and fluorescence spectroscopy (Bilge, 2020), nuclear magnetic resonance spectroscopy (Okaru et al., 2020) and inductively coupled plasma-optical emission spectroscopy (Endaye et

al., 2020). All of these methods are expensive and time-consuming. They also make sample preparation for large sample sets nearly difficult. To enable quality control management and enhance socioeconomic advantage, a quick and easy analytical method of identification, differentiation, and fraud detection concerning the geographical origin of coffee beans is required.

Handheld near-infrared spectroscopy is an analytical technique that has been found useful for quantitative and qualitative analyses in industries such as agricultural and food industries, pharmaceuticals, textiles, petrochemical, and medicine. When comparing the benefits and disadvantages of handheld NIR, the instrument's primary drawback is the instrument's limited window for spectrum acquisition (from 740nm-1070nm), whereas its advantages are generally related to the movement of the device from field to field to measure, examine and regulate the products. This analytical tool has advantages over the traditional analytical methods. The merits are; a rapid technique that requires minimal sample preparation, accurate, environmentally friendly, non-invasive, semi/non-destructive and allows simultaneous analysis. NIR spectroscopy has successfully been used for the geographical differentiation of coffee beans (Giraudo et al., 2019; Marquetti et al., 2016).

Other uses of near-infrared spectroscopy are: Identification of coffee leaves (Mees et al., 2018), prediction of antioxidants in roasted and spent coffee (Catelani et al., 2017b; Páscoa et al., 2013), coffee adulteration (Ryckewaert et al., 2020), coffee quality (Baqueta et al., 2021) and prediction of caffeine in coffee (Budiastra et al., 2018). NIR instrument in almost the same range have been carried out by

other researchers such as (Pahlawan & Masithoh, 2022) who combined Vis-NIR spectroscopy and PLS-DA for classification of Arabica and Robusta roasted coffee beans with a wavelength of 450-950nm. Dharmawan et al. (2023) developed PCA-MLP model based on visible and shortwave NIR spectroscopy for authenticating the origin of Arabica coffee. Portable NIR spectroscopy was also used to detect and quantify coffee husk in coffee (Boadu et al, 2023) and authenticate coffee bean varieties (Boadu et al., 2022) with spectra wavelength range of 740-1070 nm.

Discrimination models are built employing classification algorithm, which are supervised pattern recognition techniques. In the course of the analysis, selecting the appropriate type is crucial. Partial least square-discrimination analysis (PLS-DA) seeks to build models that can enhance separation among classes of objects (Suhandy & Yulia, 2017). Support vector machine (SVM) increases the classifier's flexibility and reduces the burden on the experimenter to develop a separate process for removing outliners as classes of data are developed (Doran et al., 2007). Linear discrimination analysis (LDA) minimizes the variance within groupings and maximizes variances between grouping (Yu et al., 2009) in any specific data set, thereby ensuring the utmost separability (Mondel & Das, 2014). Neural network (NN) is applied in optimization methods, intrusion detection and data classification. It is an excellent identifier of trends in data and patterns suited for forecasting and prediction needs (Abiodun et al. 2008). Random forest (RF) provides a way to detect outliners by using proximity analysis which can handle categorical data, unbalanced data and as well as data with missing values (Pal, 2005).

The coffee industry has recently demonstrated interest in calibration models for quantitative and qualitative research of different coffee types (bean, roasted, powdered), employing chemometric techniques in combination with several methods (Barbin et al., 2014). However, most of the applications focus on the geographical differentiation of coffee and individual assessment of coffee species (arabica and Robusta) and no studies were found in the geographical differentiation of African coffee types (bean, roasted, powdered) using the handheld near-infrared spectroscopic method.

4.2 Materials and Methods

4.2.1 Coffee sample preparation

A total of 190 samples of Robusta coffee beans were collected from five African countries- Ghana (40), Ivory Coast (40), Nigeria (40), Uganda (40) and Burkina Faso (30). These samples were prepared into three forms (raw beans, roasted, powdered) hence a total of 570 samples. They were collected from October to December, 2021. The coffee beans were thoroughly dried, without any unusual foreign odours or signs of adulteration. They were uniform in size, free of living insects, broken beans, shell fragments, and foreign materials. The coffee samples were sent to the School of Agriculture Laboratory at the University of Cape Coast in zip-locked polyethene bags for further investigation.

4.2.1 Coffee bean roasting and grinding

Each sample of coffee collected from the five countries was roasted separately for one hour at a temperature of 200 °C (medium roast) according to

Vasconcelos et al. (2007). Coffee beans were roasted at the same temperature and power for purposes of comparison. A home electric coffee bean roaster (JIAWANSHUN, China) that is capable of holding 750 grams of coffee at a time was used to roast the coffee. The same quantities of roasted Robusta coffee beans from each of the five different countries were processed individually into a powder using a multipurpose grinder (QE-100, Zhejiang YiLi TTool Co, Ltd, China). The powder coffee was sieved with a mesh size of 200 um. For subsequent analysis, all samples were kept in zip-lock bags.

4.2.2 Reference measurement of coffee bean quality

In this investigation, the raw, roasted, and powdered states of coffee beans were examined using wet chemistry analysis of the quality parameters. The coffee samples were scanned and immediately analysed using standard wet chemistry reference methods as described in (AOAC, 1990). Parameters such as moisture content, lipids, carbohydrates, proteins, and ash were measured with three independent replications. Moisture was determined by oven-drying method at 105^{o}C. Ash content was determined by incineration of the sample at 550^{o}C in a furnace. Total nitrogen content was analysed by Kjeldahl procedure. Protein was calculated from total nitrogen using a factor of 6.25. Lipid was measured by the Soxhlet method by extraction with diethyl ether solvent. Dietary fibre was carried out by fibertec system and carbohydrate was calculated by difference.

4.2.3 Statistical analysis

The means and standard deviation of all three replications were analysed using Minitab 16 software. The differences among the test parameters were identified by one-way analysis of variance (ANOVA) using Fisher's least significance differences (LSD) test. All statistical tests were carried out at a 5% significance level.

4.2.4 Spectra acquisition using handheld NIR spectroscopy

The scanning of raw coffee beans, roasted coffee beans, and powdered samples kept in clear zip-locked polyethene bags from various locations was done using a handheld NIR spectrometer (Scio, UK) controlled by a smartphone (Samsung A21). The wavelength range of the NIR spectral data was 740-1070 nm, with a resolution of 1 nm. The portable spectrometer reports the measurement in units of relative abundance (log 1/R) that were connected to chemical ingredients. Five scans were performed on each coffee sample and the average was taken. There was no noticeable interference with the NIR signals from the transparent zip-locked bags.

4.2.4.1 Spectral data partition

The spectral dataset for raw coffee beans (190), roasted coffee beans (190), and powdered coffee (190) were all downloaded independently, and each category was divided into two subsets: a training set and a prediction set. A total of 126 samples and 63 samples were selected as the training set and testing set, respectively, for the raw coffee beans, roasted coffee beans, and powdered coffee.

The model was developed using the training data, and its actual prediction was assessed using the testing data. As done by previous writers, the individual samples in each set were randomly selected to form a 2/1 partition of the training set and testing set (Anyidoho et al., 2020; Teye et al., 2014).

4.2.4.2 Software device, spectrum preprocessing methods, PLS-DA

MATLAB (Math Works, Inc, USA, version 9.6.0) with Windows 10 Pro software package was used for computation, chemometric analysis, and generation of figures. Five replicates were collected from the NIR spectra of each coffee sample and states were averaged. Chemometric methods; principal component analysis (PCA) and mathematical preprocessing methods were investigated to eliminate errors and reduce the dimensions while maintaining the similarities and differences between observations as much as possible. The first derivative (FD), second derivative (SD), mean centering (MC), Multiple scattering correction (MSC), as well as Standard Normal Variate (SNV), were applied in the study. These techniques have been used for classification problems by other authors (Anyidoho et al., 2020; Marquetti et al., 2016). It was developed for each spectrum (FD, SD, MC, MSC, SNV) from each coffee bean state to discriminate Robusta coffee samples from different countries in Africa. PCA as a non-supervised pattern recognition tool was applied to the spectra to observe a possible cluster trend that differentiates coffee types (bean, roasted, and powdered) from different countries. Partial least squares discriminant analysis (PLS-DA) algorithm was used for building the classification model for detecting coffee samples from different countries.

4.2.5 Theory of preprocessing and modelling techniques

There are numerous baseline removal techniques, including spectral derivative transformation, one of the most effective techniques for minimizing baseline defects. (Duckworth, 2004): Baseline offset may be effectively removed using the first derivative (FD), and baseline drifts and the linear trend can be removed from a spectrum using the second derivative (SD) preprocessing techniques (Duckworth, 2004; Rinnan et al., 2009) overlapping peaks, while improving minute spectral differences, and detaching peaks. The mean centering (MC) preprocessing technique is performed by determining the average spectrum of the data set and deducting the average from each spectrum. The multiple scattering correction (MSC) method corrects the addictive and multiplicative effects that occur due to different particle sizes and orientations as well as morphology (Dhanoa et al., 1994). It also prevents scattered light of different particle sizes. According to Barnes et al. (1989), the standard normal variate (SNV) eliminates slope changes when objects are changed independently as well as additive and multiplicative scatter effects (Candolfi et al., 1999).

4.2.6 Principal component analysis (PCA)

This is used to analyse data obtained from modern measurement techniques and neatly organized in the data matrix (Massart et al., 1998). It is a technique for reducing the original data space's dimensionality by employing a smaller and more efficient abstract space of latent variables, where the data can be displayed and the information from the original space is essentially reserved. The information is reduced into new variables called PCs during the reduction of the data matrix,

where PC1, PC2, and PC3 typically offer and explain the most important information in descending order.

4.2.7 Multivariant classification algorithm

After the application of spectral preprocessing methods, a variety of multivariate classification algorithms were systematically examined, and the outcomes were compared. These classifications include the Partial Least Squares - Discriminate Analysis (PLS-DA) algorithm, the Support Vector Machine (SVM), the Linear Discriminant Analysis (LDA), the Neural Network (NN), and the Random Forest (RF). For more information on the theories, reference other authors. (Cao et al., 2012; Christin et al., 2013; Thapngam et al., 2012; Thibault et al., 1990; Yu & Kim, 2012).

4.2.8 Model validation

The classification models were evaluated using K-fold cross validation (Marcot & Hanea, 2021). Furthermore, accuracy, precision, recall, and F1-score were also used to assess the performance of the different classification models. True positive (TP), true negative (TN), false positive (FP), and false negative (FN) abbreviations were used in equations 1 through 4.

$$\text{Accuracy} = \frac{TN+TP}{TN+TP+FN+FP} \quad (4.1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (4.2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4.3)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision x Recall}}{\text{Precision+Recall}} \quad (4.4)$$

4.3 Results and Discussion

The proximate compositions of coffee beans in five African countries have been shown in Table 4.7. Moisture content is an important parameter for the evaluation of green coffee quality because it causes the growth of moulds, production of mycotoxins, and affects the sensory, chemical, and physical parameters. The moisture content for the five African countries for raw Robusta coffee was in the range of 8.24 to 12.67%. They were lower than the values (12 to 13 %) reported by the literature for Robusta coffee beans (Franca & Oliveira, 2008). This could prevent the beans from deteriorating during transportation and storage (Reh et al., 2006). The ashes recorded for the five African Robusta coffee beans were in the range of 3.09 to 4.23%, slightly lower than the values reported in the literature while that of protein (14.41 to 16.76) was slightly higher. Lipid (3.73 to 7.09%) and carbohydrate values (67.47 to 70.81%) were slightly lower than the values reported in the literature for raw coffee beans (Franca & Oliveira, 2008).

Table 4.7: Proximate composition of Robusta coffee beans in five African countries

Sample	Moisture (%)	Ash (%)	Protein (%)	Lipid (%)	Fibre (%)	CHO (%)
Ghana	9.39±0.07 ^{a}	3.33±0.18 ^{bc}	16.40±0.10 ^{a}	5.67±0.18 ^{a}	7.13±0.13 ^{a}	67.47±0.20 ^{d}
Ivory Coast	10.89±0.07 ^{b}	3.09±0.06 ^{c}	15.03±0.10 ^{b}	3.73±0.13 ^{b}	7.35±0.28 ^{a}	70.81±0.48 ^{a}
Nigeria	10.28±0.14 ^{c}	3.39±0.21 ^{b}	14.41±0.21 ^{c}	7.09±0.11 ^{c}	6.54±0.31 ^{b}	68.56±0.34 ^{c}
Burkina Faso	12.67±0.12 ^{d}	3.61±0.15 ^{b}	16.51±0.33 ^{a}	4.55±0.28 ^{b}	5.99±0.13 ^{c}	69.34±0.50 ^{bc}
Uganda	8.24±0.15 ^{e}	4.23±0.15 ^{a}	14.76±0.21 ^{bc}	4.13±1.19 ^{b}	7.24±0.23 ^{a}	69.65±0.10 ^{b}

Means in columns that do not share a letter are significantly different $p < 0.05$
CHO- Carbohydrate

The proximate compositions of roasted coffee for the five African countries are shown in Table 4.8. The ashes were within the range (3.50 to 5.53%), protein was slightly higher ranging from 16.03 to 18.73%, lipids (6.30 to 10.33%), and carbohydrates (62.12 to 66.90%) were slightly lower as reported in literature (Franca & Oliveira, 2008; Gyedu-Akoto et al., 2019a). Comparing the proximate composition of raw and roasted coffee, there were no significant variations except for moisture content.

Table 4.8: Proximate composition of roasted Robusta coffee in five African countries

Sample	Moisture (%)	Ash (%)	Protein (%)	Lipid (%)	Fibre (%)	CHO (%)
Ghana	4.92±0.24 ^{ab}	4.58±0.19 ^{a}	16.14±0.19 ^{a}	7.69±0.13 ^{b}	5.47±0.16 ^{ab}	66.12±0.43 ^{bc}
Ivory Coast	4.41±0.20 ^{b}	3.50±0.29 ^{b}	17.99±0.17 ^{b}	6.30±0.26 ^{d}	5.31±0.04 ^{b}	66.90±0.64 ^{a}
Nigeria	4.80±0.01 ^{ab}	5.53±0.32 ^{c}	16.03±0.01 ^{a}	10.33±0.20 ^{a}	5.77±0.26 ^{a}	62.35±0.16 ^{d}
Burkina Faso	5.04±0.30 ^{a}	3.69±0.52 ^{b}	18.73±0.33 ^{c}	6.68±0.21 ^{cd}	5.32±0.17 ^{b}	65.58±0.30 ^{c}
Uganda	5.23±0.46 ^{a}	4.94±0.09 ^{a}	16.21±0.05 ^{a}	6.75±0.27 ^{c}	5.38±0.14 ^{b}	66.71±0.28 ^{ab}

Means in columns that do not share a letter are significantly different $p < 0.05$
CHO- Carbohydrate

4.3.1 Spectroscopic data presentation for coffee bean, roasted and powdered

The spectra of coffee show unique fingerprints according to their groupings. Figure 4.10 illustrates the raw and mean spectra of samples of Robusta coffee beans from the five African countries. Each spectrum can be seen to have a distinctive profile because of the difference in the chemical composition as shown in Table 4.7. For near-infrared spectroscopy, electromagnetic radiation absorption is based on a wavelength range of 780–2500 nm. This study's absorbance was measured

between 740 and 1070 nm in wavelength. The obtained spectra from the NIR scanning of the Robusta coffee samples revealed several bands and a single peak, as seen in the profile provided. These bands are composed of overtones and combinations of fundamental vibrations that are related to their biological and biochemical properties. The figure shows that each profile has a similar profile, as can be seen in Figure 4.10. Again, the closer and overlapping spectra as discovered in the raw and mean spectra are countries that share a common boundary with the other. Specifically, in Figure 4.10 (iv) at the wavelength of 750 nm - 850 nm, Ghana and Ivory Coast spectra were closer for the raw coffee. Due to their shared border and proximity to one another in terms of where they reside on the African continent, Ghana and the Ivory Coast displayed comparable patterns. The spectra for Ghana and Uganda at wavelength 1000 nm - 1050 nm were very close which could be due to similarities in biochemical and organic properties. There was a wide gap between the spectra of Nigeria coffee and the spectra of other countries for raw, roasted, and powdered coffee as shown in Figure 4.10 (iv, v, and vi). It was revealed by this phenomenon that locations could be identified. Furthermore, it implies that when two regions share a border, similar farming practices (pre- and post-harvest activities) that affect coffee bean quality features are not wholly different. Additionally, the wavelengths of 750 nm to 900 nm are associated with the third overtone of the C-H stretching vibration, which stands for carbohydrates, proteins, and lipids, while the wavelengths of 900 nm to 1050 nm are connected to the second overtone of the N-H stretching, which stands for fat and proteins, respectively (Barbin et al., 2014). Different amounts of protein and lipid content in the samples

resulted in spectral variations in the NIR spectra (Ryan et al., 2006). Roasting and grinding of coffee bring about differences in their nutrients which results in spectra variations as revealed in Figure 4.10 for bean, roasted and powder coffee samples.

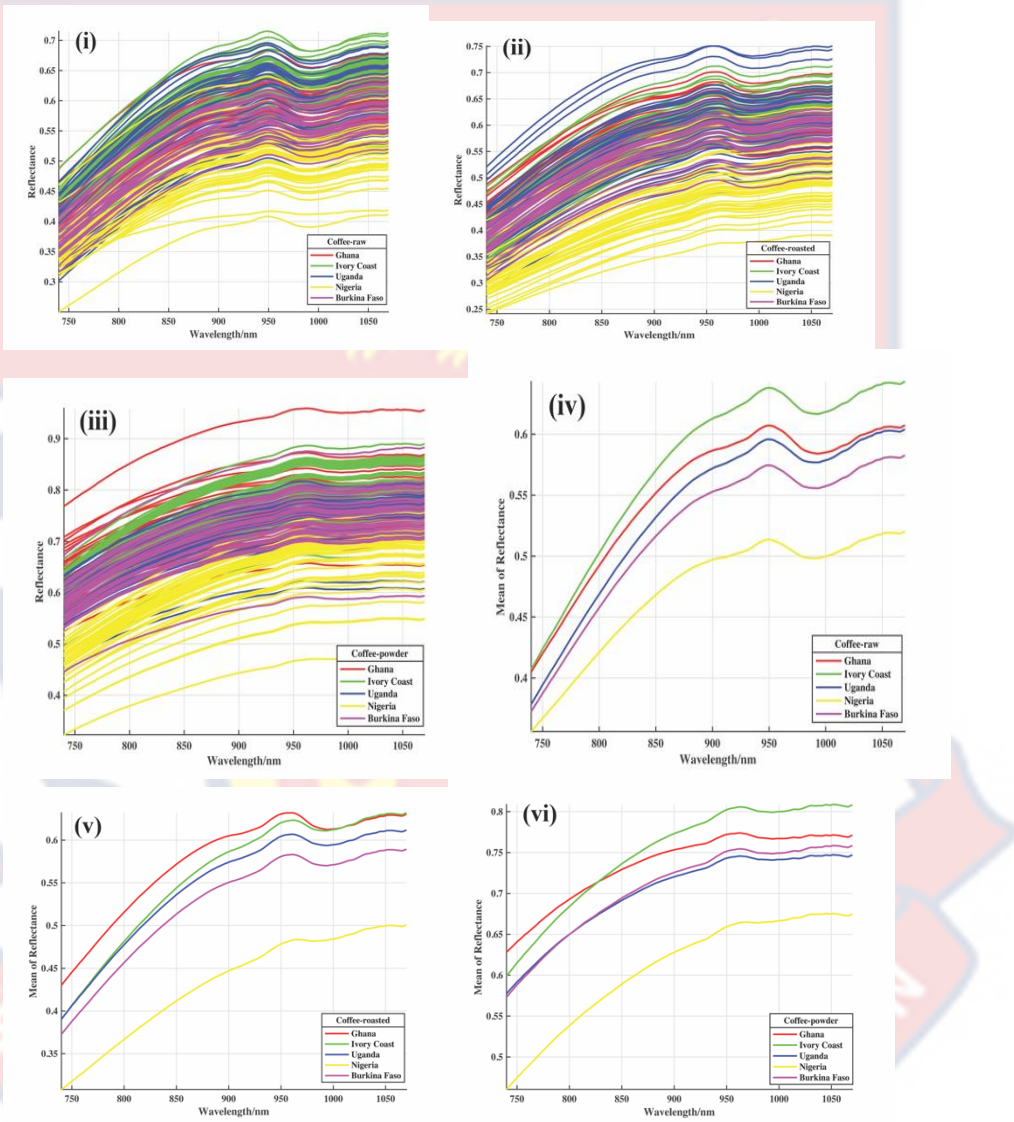


Figure 4.10: Spectra of Robusta coffee types: (i) raw bean, (ii) roasted bean, (iii) powder and mean spectra for coffee types: (iv) raw bean, (v) roasted bean, (vi) powder from five African countries

Some chemical substances like water and protein are lost when heat is applied to them. Based on a trial-and-error approach, several preprocessing

treatments were selected to enhance the classification model. The classification of Robusta coffee for African countries also utilized the PLS-DA model.

4.3.2 Mean Spectra representation for coffee raw bean, roasted bean and powdered coffee

The mean spectra representation for raw, roasted and powdered samples exhibit absorption band in the regions of 940 – 1000 nm corresponding to the third overtone as seen in Figure 4.11. When compared to raw (green) and roasted coffee, the band in the 940–1000nm range that corresponds to the third overtone of CH, CH_2, and CH_3 groups was more prevalent and defined in the roasted powder samples. Furthermore, the powdered samples show wide spectra differences. This could be as a results of the smaller particle size in the powdered samples which are likely to allow easy penetration and cause an increased absorption because lower absorbance is found in coffee samples with larger particle sizes, and vice versa. (Shan et al., 2014).

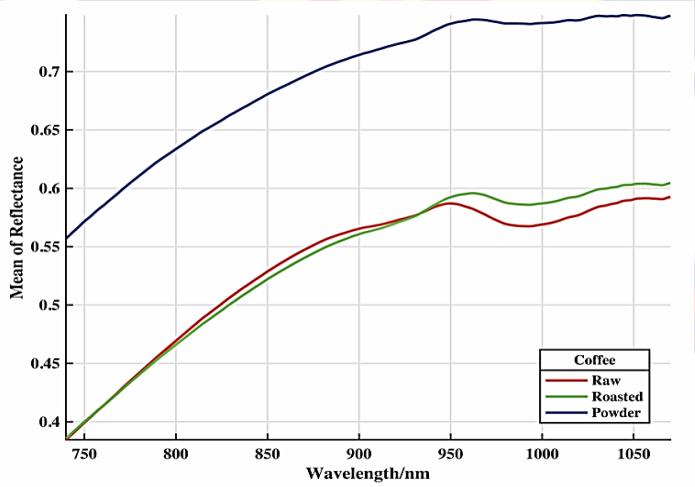


Figure 4.11: Raw mean spectra for coffee types: (a) raw bean, (b) roasted bean, (c) powdered coffee

PCA for African countries

4.3.2.1 Green coffee bean

Robusta coffee bean sample spectra from five different countries were analyzed with PCA and classified with PLS-DA. PCA was able to provide spectral data and trends but was not a classification tool like PLS-DA. The PCA was used to classify the five groups of Robusta coffee into three principal components (PC1, PC2, and PC3) because they displayed a separate distinction among all the attributes examined. The score plot of PCA for the pre-processed techniques of Robusta bean for the five African countries is shown in Figure 4.12.

MC provides the most effective treatment effect in PCA with 99.76%, followed by MSC with 99.18%, SNV with 99.06%, FD with 96.85%, and SD with 74.35%. The three topmost PCs for MC as shown in Figure 4.12 (iv) contributed to 99.79% (PC1 = 87.57, PC2 = 11.90, PC3 = 0.29) of the total variance of the dataset for the coffee beans of five African countries. This indicates that the top three PCs of MC, which encompass the essential chemical fingerprint of Robusta coffee samples, can explain 99.77% of the variance information from the spectral data.

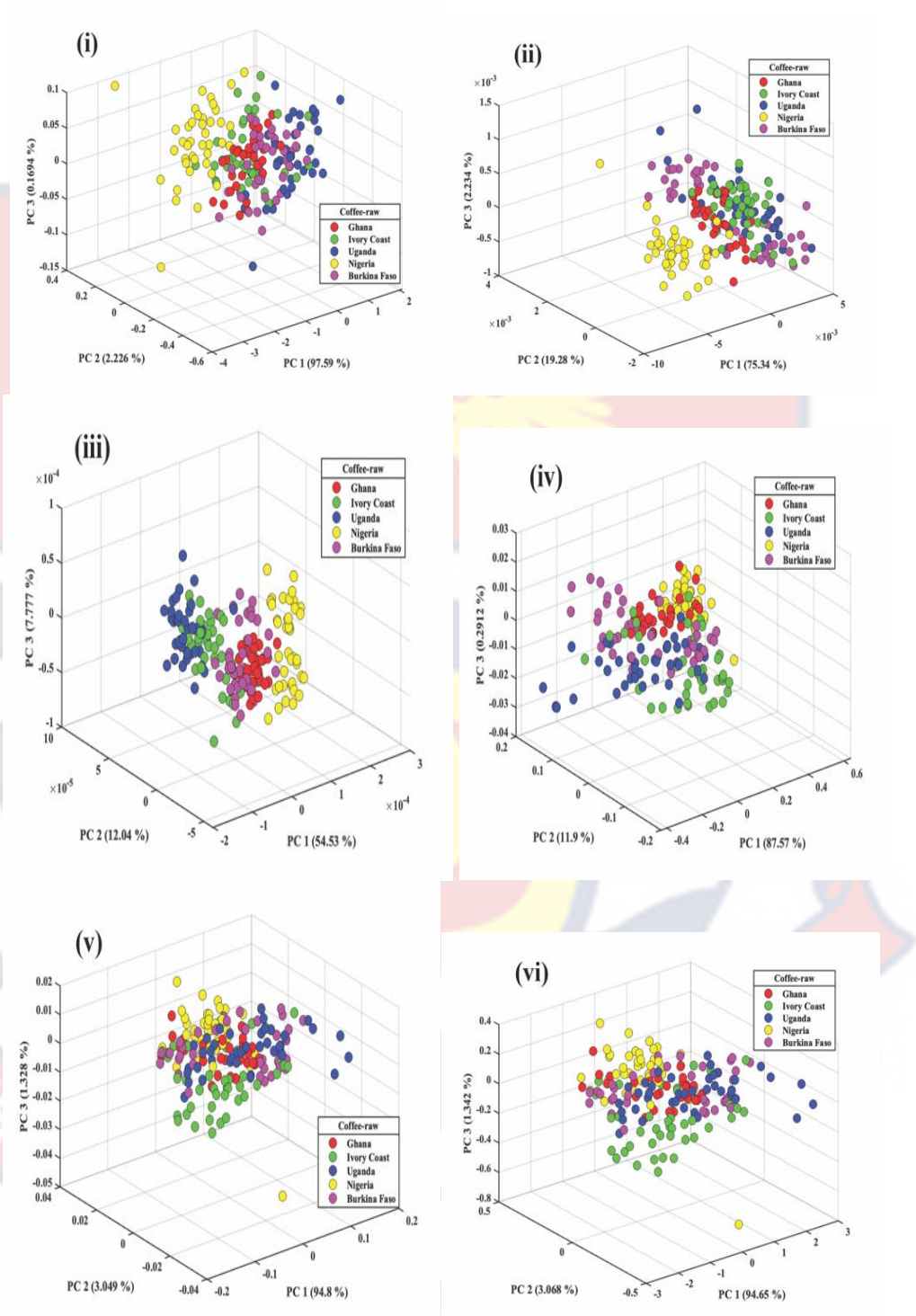


Figure 4.12: PCA score plot of the first three PCs of raw coffee beans from five African countries preprocessed- i) raw ii) FD iii) SD iv) MC v) MSC vi) SNV

4.3.2.2 Roasted coffee bean

The PCA for preprocessed techniques for roasted coffee beans of the five African countries is shown in Figure 4.13.

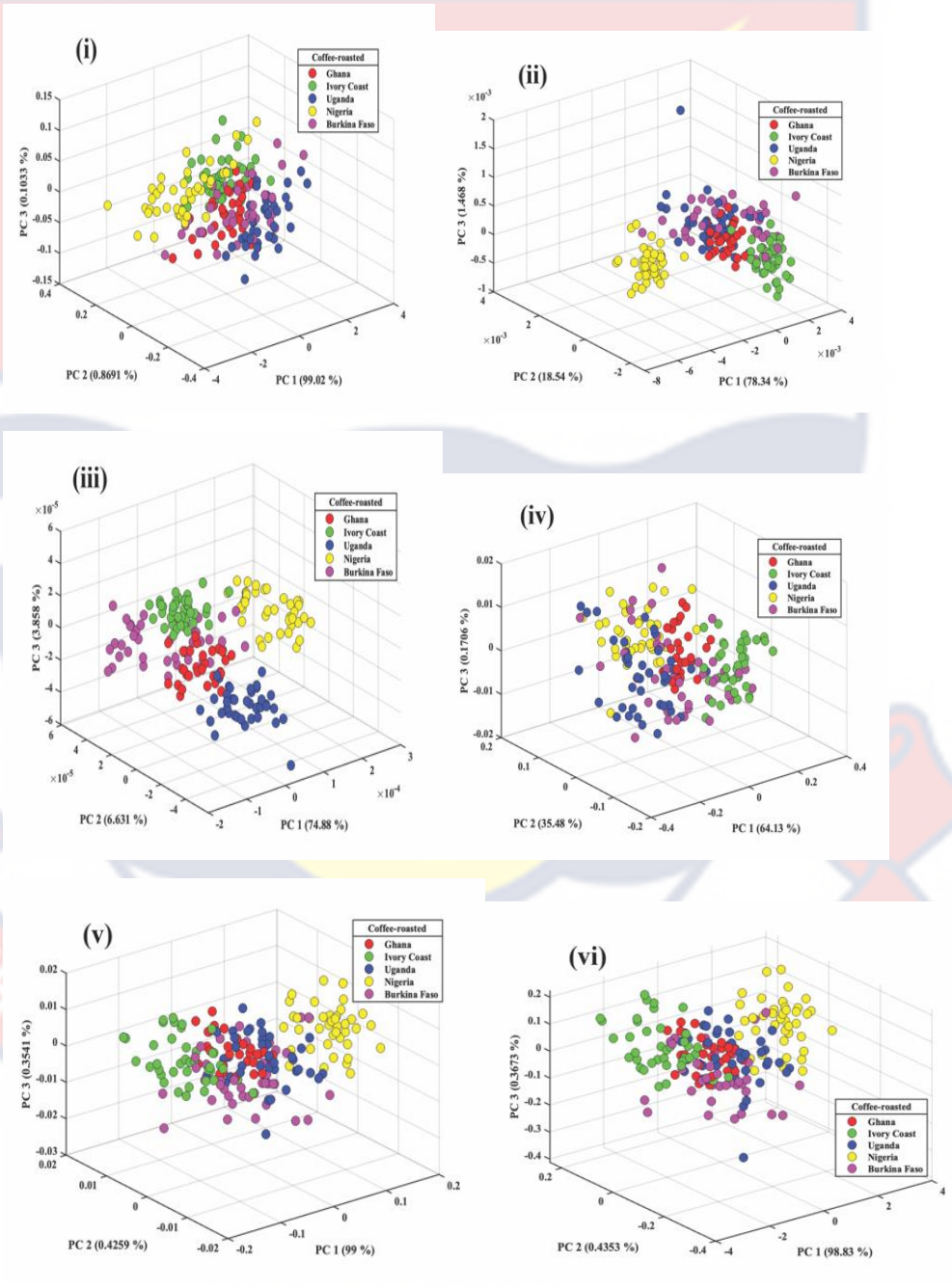


Figure 4.13: PCA score plot of the first three of roasted coffee beans from five African countries preprocessed - i) raw ii) FD iii) SD iv) MC v) MSC vi) SNV

MC and MSC performed better than the other preprocessed techniques by scoring 99.78%. Although they have equal performance, PC1 of MSC was able to classify 99% of the roasted coffee beans while PC1 of MC was able to classify only 64.13%. PC1 of SNV was able to classify 98.83%, but the three PCs scored a lower mark (99.63%) compared to the three PCs of MC and MSC. The three topmost PCs of FD were able to classify 98.35% and SD classified 85.37%. As a result, 99.78% of the variance information from the spectral data can be explained by the top three PCs of MC and MSC.

4.3.2.3 Roasted coffee powder

The PCA of the roasted coffee powder for preprocessed techniques of the five African countries is shown in Figure 4.14. MC performed better than the other pre-processed techniques by scoring 99.88%, MSC with 98.97%, SNV with 98.89%, FD with 96.81% and SD with 58.83% in the descending order. As shown in Figure 4.14(iv), the top three PCs for MC accounted for 99.88% of the variation in the dataset for roasted coffee powder (PC1 = 69.97%, PC2 = 2.77%, and PC3 = 0.14%). This indicates that 99.88% of the variance information from the spectral data can be explained by the top three PCs of MC.

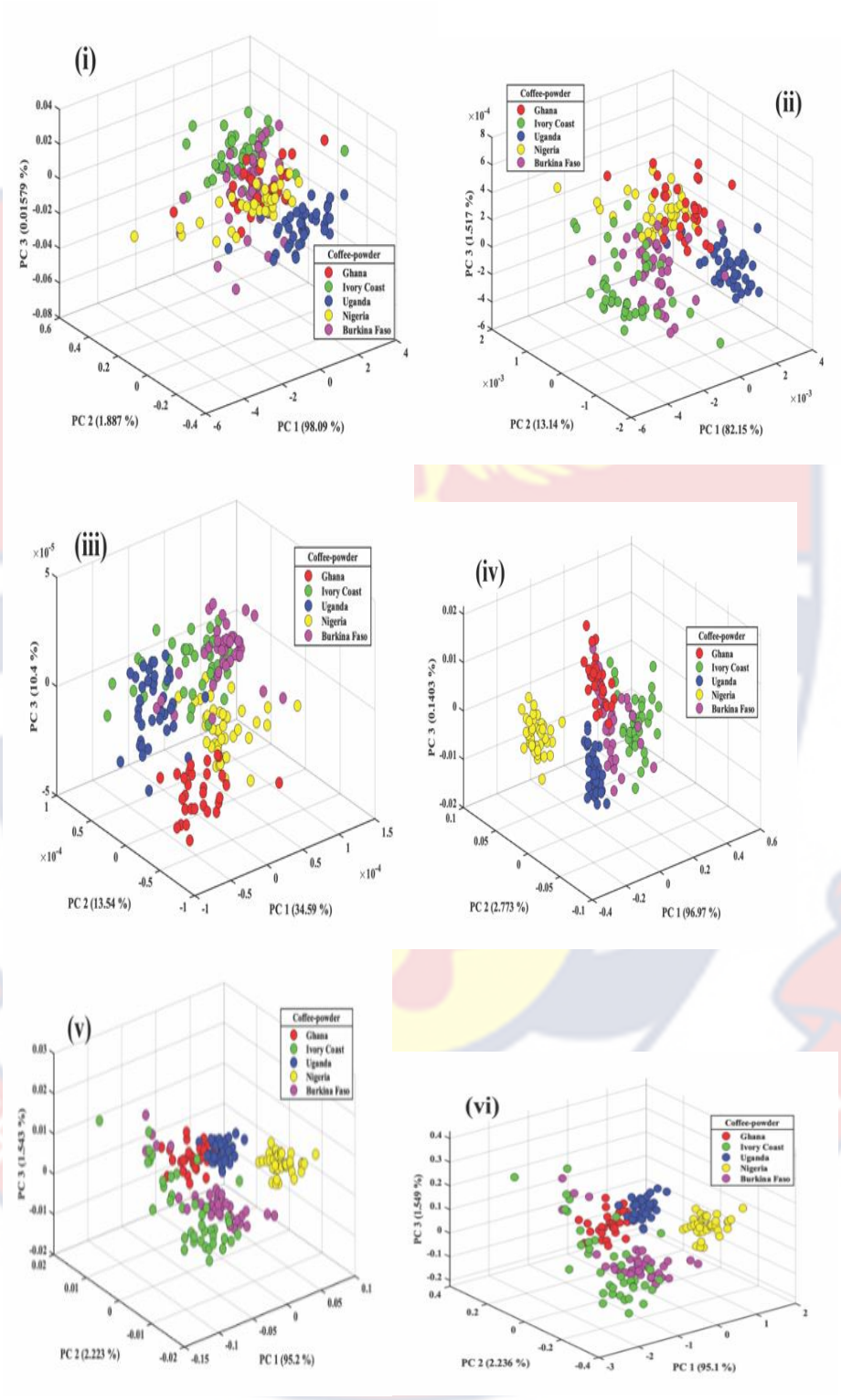


Figure 4.14: PCA score plot of the first three of roasted coffee powder from five African countries preprocessed- - i) raw ii) FD iii) SD iv) MC v) MSC vi) SNV

4.3.3 PLS-DA of Robusta from African countries

4.3.3.1 Green coffee bean

As illustrated in Figure 4.15, where a distinction between the Robusta coffee samples by country can be observed, a PLS-DA model was developed for each spectrum preprocess. With accuracy and an FI-score of 0.87 in the classification, the score plot of the PLS-DA model built with SD in Figure 4.15 (iii) performed better in the separation. While Uganda and Ivory Coast samples were differentiated by the negative part of both the x and y axes, Burkina Faso samples were distinguished by the negative part of the x-axis and around the zero (neutral) region of the y-axis. The negative portion of the x and y axes was employed to separate the Ghana samples.

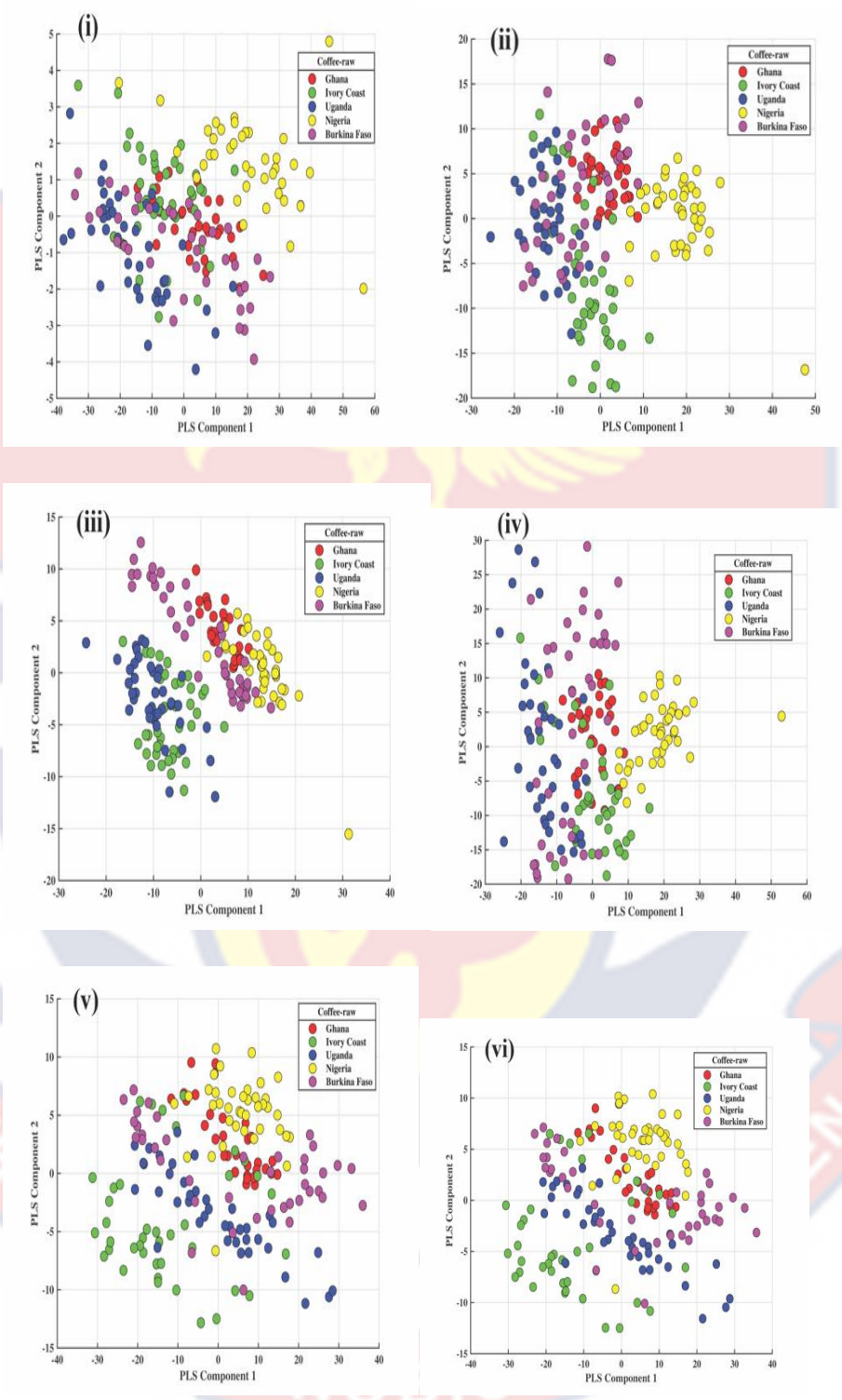


Figure 4.15: Score plot of PLS-DA model built from spectra of raw coffee bean from five African countries: i) raw ii) FD iii) SD iv) MC v) MSC vi) SNV for Robusta

4.3.3.2 Roasted coffee bean

The results obtained as shown in Figure 4.16 display the PLS-DA preprocessing techniques used for the roasted coffee beans of five countries in Africa. Due to its high F1 score and accuracy of 0.91, the score plot for PLS-DA with SD is regarded as the best for classifying roasted coffee beans. Figure 4.16 (iii) shows the separation of the samples, showing that the samples from Nigeria were distinguished by the positive part of the x-axis and the negative part of the y-axis, while the samples from Ghana, Uganda, and Burkina Faso were distinguished by the positive part of the x-axis and the range of -10 to +10 on the y-axis. Both the negative x and y axes were used to distinguish Ivory Coast samples.

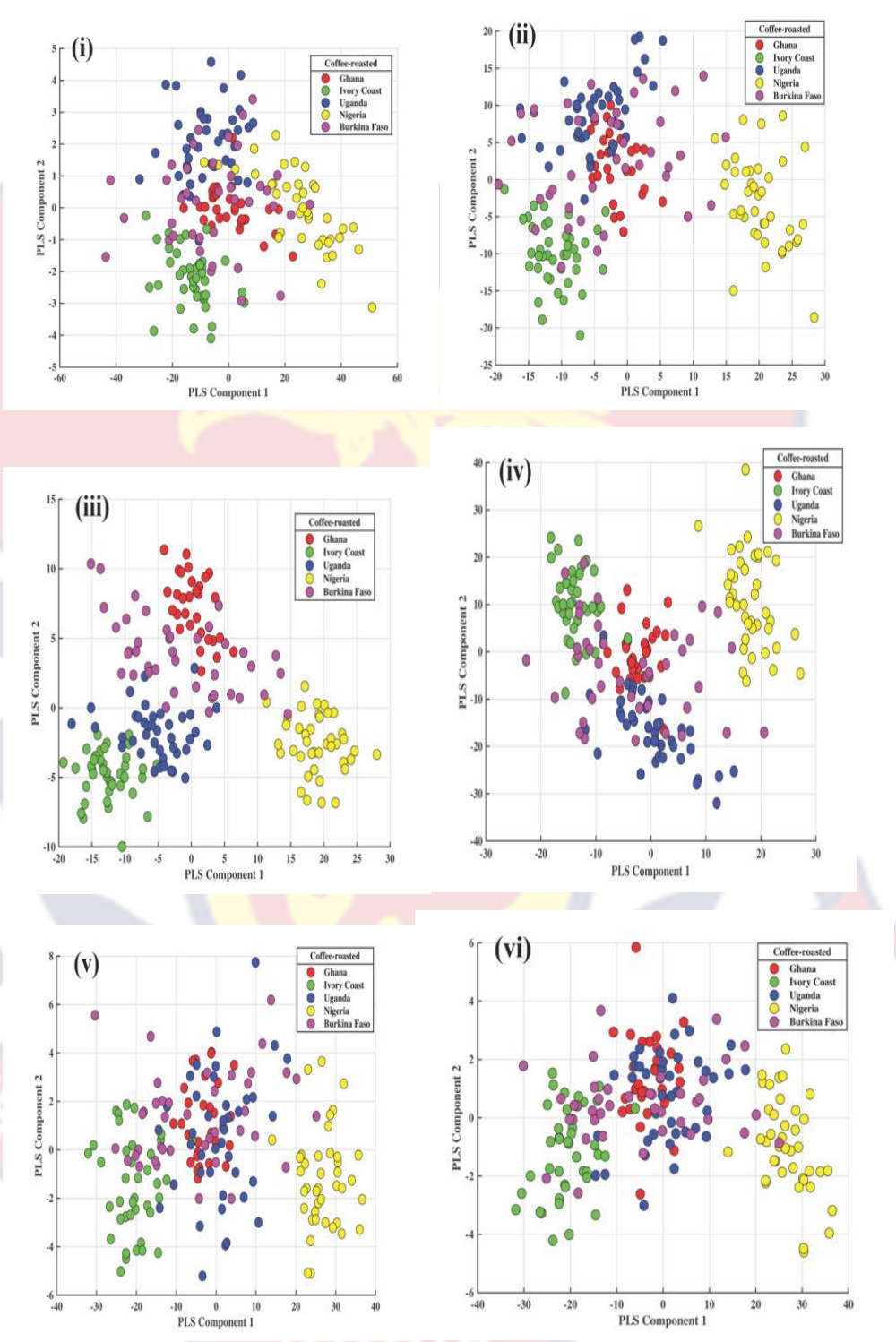


Figure 4.16: Score plot of PLS-DA model built from spectra of roasted coffee from five African countries: i) raw ii) FD iii) SD iv) MC v) MSC vi) SNV for Robusta

4.3.3.3 Roasted coffee powder

PLS-DA preprocessing methods for roasted coffee powder from five African countries are presented in Figure 4.17. Due to its high accuracy of 0.92 and F1 score of 0.93, the score plot for PLS-DA with MSC is regarded as the best in the classification of roasted coffee powder. Figure 4.17 (v) illustrates a division showing that samples from Nigeria were distinguished by both the positive and y-axis, whereas samples from the Ivory Coast were distinguished by both the negative and y-axis. Uganda was differentiated based on the positive and negative axes of the y-axis.

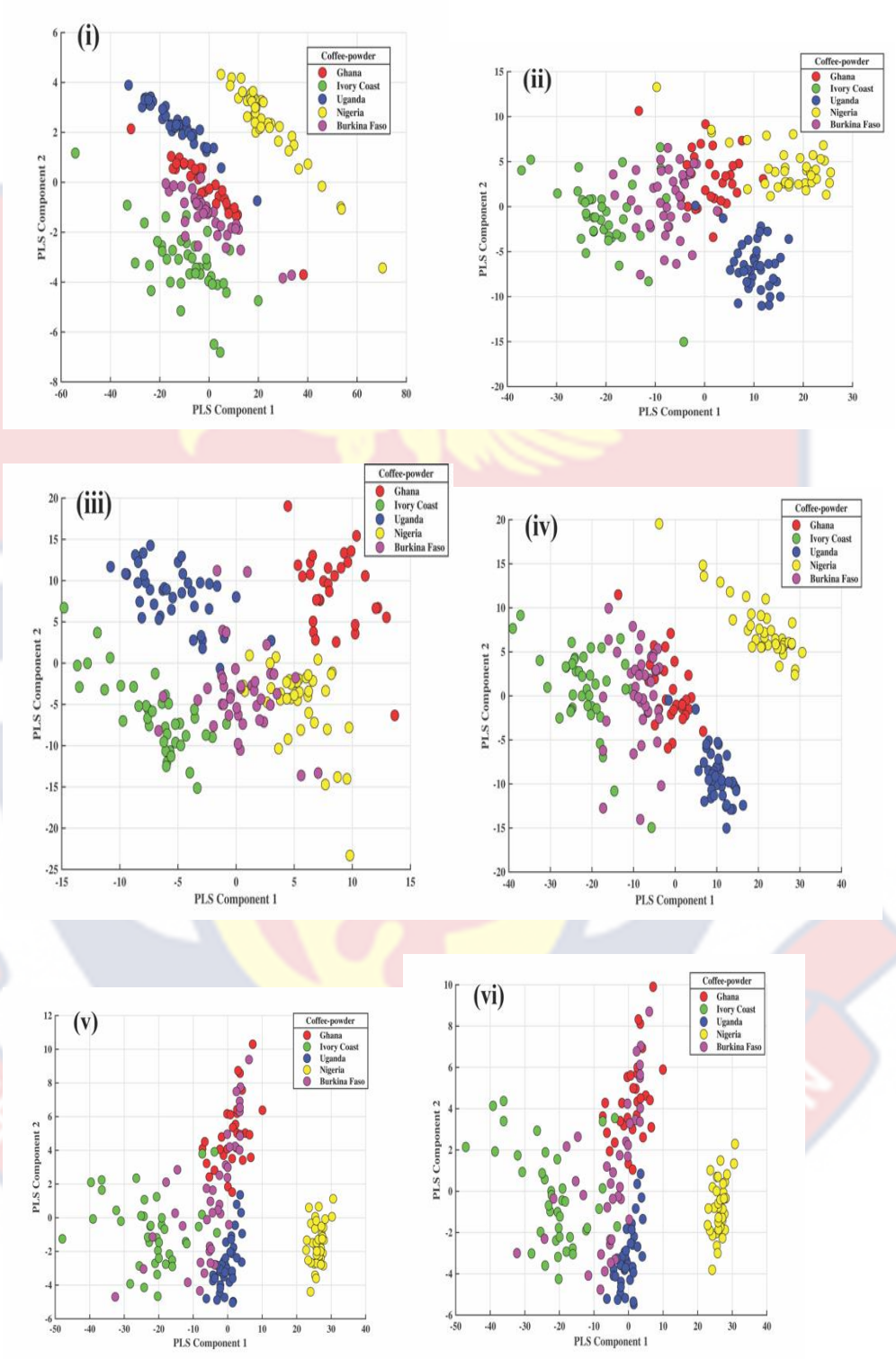


Figure 4.17: Score plot of PLS-DA model built from spectra of roasted coffee powder from five African countries: i) raw ii) FD iii) SD iv) MC v) MSC vi) SNV for Robusta

4.3.4 Performance of classification model for bean, roasted and powdered coffee

To search for a better classification for the coffee bean, SVM, LDA, NN, RF, and PLSDA models were developed with a single or combination of two different preprocessing techniques. As done by previous writers, accuracy and F1-score were used to assess the model's performance (Gorji et al., 2022; Priyadarshini et al., 2018). The harmonics mean between recall and precision is the F1-score. (Dalianis, 2018). Table 4.9 shows the classification results in summary. Among the classification models, it was observed that the second derivative with PLSDA and MC plus the second derivative with PLSDA performed better with an accuracy of 0.87 and an FI-score of 0.88. SNV plus the second derivative provided the best SVM model for the roasted coffee samples, as shown in Table 4.10, with an accuracy of 0.97 and an F1-score of 0.96. According to Table 4.11, MSC plus second derivative was the most accurate NN model, with an F1-score of 0.96. In comparison to previous models, the best reported values for all types of coffee were all nearer to one (1). Even though there were no appreciable differences among the parameters examined, the handheld NIR spectroscopy was able to geographically differentiate the various types of African coffee (bean, roasted, and powdered).

Table 4.9: Performance of classification algorithms with single and combination of two different preprocessing techniques of coffee bean

Model		RAW		MC		MSC		SNV		SD		MC+SD		MSC+SD		SNV+SD	
		Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
SVM	Accuracy	0.64	0.64	0.69	0.69	0.57	0.57	0.57	0.57	0.71	0.71	0.71	0.71	0.69	0.69	0.70	0.70
	Precision	0.63	0.62	0.7	0.7	0.57	0.57	0.57	0.56	0.71	0.73	0.71	0.73	0.7	0.7	0.70	0.70
	Recall	0.63	0.64	0.69	0.68	0.55	0.54	0.55	0.54	0.71	0.72	0.71	0.72	0.69	0.71	0.69	0.71
	F1-score	0.63	0.63	0.7	0.69	0.56	0.55	0.56	0.55	0.71	0.73	0.71	0.73	0.69	0.7	0.7	0.71
LDA	Accuracy	0.64	0.64	0.67	0.67	0.53	0.53	0.54	0.54	0.71	0.71	0.71	0.71	0.67	0.67	0.67	0.67
	Precision	0.65	0.66	0.67	0.66	0.49	0.49	0.5	0.5	0.7	0.71	0.7	0.71	0.63	0.63	0.65	0.65
	Recall	0.63	0.62	0.67	0.67	0.51	0.5	0.52	0.51	0.7	0.71	0.7	0.71	0.64	0.63	0.65	0.63
	F1-score	0.64	0.64	0.67	0.67	0.5	0.49	0.51	0.5	0.7	0.71	0.7	0.71	0.64	0.64	0.65	0.65
NN	Accuracy	0.64	0.64	0.74	0.74	0.66	0.66	0.62	0.62	0.74	0.74	0.77	0.77	0.77	0.77	0.82	0.82
	Precision	0.64	0.66	0.75	0.75	0.67	0.69	0.62	0.62	0.75	0.73	0.77	0.78	0.77	0.79	0.81	0.81
	Recall	0.64	0.66	0.74	0.76	0.67	0.67	0.61	0.63	0.74	0.74	0.77	0.78	0.76	0.79	0.81	0.82
	F1-score	0.64	0.66	0.74	0.75	0.67	0.68	0.61	0.62	0.74	0.74	0.77	0.78	0.77	0.79	0.81	0.82
RF	Accuracy	0.63	0.63	0.76	0.76	0.59	0.59	0.66	0.66	0.79	0.79	0.79	0.79	0.85	0.85	0.82	0.82
	Precision	0.63	0.63	0.76	0.76	0.59	0.61	0.67	0.68	0.79	0.76	0.79	0.76	0.85	0.86	0.82	0.82
	Recall	0.63	0.63	0.77	0.77	0.59	0.6	0.67	0.67	0.79	0.79	0.79	0.79	0.85	0.84	0.82	0.83
	F1-score	0.63	0.63	0.77	0.76	0.59	0.6	0.67	0.68	0.79	0.79	0.79	0.79	0.85	0.85	0.82	0.83
PLSDA	Accuracy	0.66	0.66	0.69	0.69	0.61	0.61	0.65	0.65	***0.87**	**0.87**	***0.87**	**0.87**	0.84	0.84	0.85	0.85
	Precision	0.65	0.66	0.69	0.68	0.59	0.6	0.63	0.66	***0.87**	**0.88**	***0.87**	**0.88**	0.84	0.83	0.86	0.85
	Recall	0.65	0.65	0.69	0.69	0.59	0.58	0.63	0.63	***0.87**	**0.87**	***0.87**	**0.87**	0.84	0.84	0.86	0.85
	F1-score	0.65	0.66	0.69	0.69	0.59	0.59	0.63	0.64	***0.87**	**0.88**	***0.87**	**0.88**	0.84	0.84	0.86	0.85

***Best model**

Table 4.10: Performance of classification algorithms with single and combination of two different preprocessing techniques of roasted coffee

Model		**RAW** Training set	Test set	**MC** Training set	Test set	**MSC** Training set	Test set	**SNV** Training set	Test set	**SD** Training set	Test set	**MC+SD** Training set	Test set	**MSC+SD** Training set	Test set	**SNV+SD** Training set	Test set
SVM	Accuracy	0.68	0.68	0.68	0.68	0.69	0.69	0.67	0.67	0.95	0.95	0.95	0.95	0.96	0.96	***0.97**	**0.97**
	Precision	0.67	0.68	0.68	0.68	0.67	0.68	0.66	0.67	0.95	0.94	0.95	0.94	0.96	0.96	***0.97**	**0.97**
	Recall	0.66	0.69	0.66	0.68	0.68	0.7	0.66	0.69	0.95	0.96	0.95	0.96	0.96	0.96	***0.97**	**0.97**
	F1-score	0.67	0.67	0.67	0.67	0.68	0.69	0.66	0.68	0.95	0.95	0.95	0.95	0.96	0.96	***0.97**	**0.97**
LDA	Accuracy	0.66	0.66	0.68	0.68	0.68	0.68	0.68	0.68	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
	Precision	0.63	0.63	0.66	0.68	0.66	0.67	0.67	0.68	0.95	0.95	0.95	0.95	0.95	0.96	0.95	0.96
	Recall	0.64	0.64	0.67	0.68	0.67	0.68	0.67	0.68	0.95	0.96	0.95	0.96	0.95	0.94	0.95	0.94
	F1-score	0.64	0.64	0.67	0.68	0.67	0.67	0.67	0.68	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
NN	Accuracy	0.70	0.70	0.69	0.69	0.61	0.61	0.62	0.62	0.93	0.93	0.64	0.64	0.64	0.64	0.94	0.94
	Precision	0.71	0.69	0.69	0.69	0.6	0.6	0.63	0.63	0.93	0.94	0.64	0.64	0.64	0.64	0.94	0.94
	Recall	0.7	0.7	0.69	0.69	0.6	0.6	0.62	0.61	0.92	0.93	0.64	0.63	0.64	0.63	0.94	0.93
	F1-score	0.7	0.69	0.69	0.69	0.6	0.6	0.62	0.62	0.93	0.94	0.64	0.63	0.64	0.63	0.94	0.94
RF	Accuracy	0.71	0.71	0.75	0.75	0.65	0.65	0.66	0.66	0.92	0.92	0.92	0.92	0.96	0.96	0.94	0.94
	Precision	0.68	0.71	0.73	0.76	0.64	0.66	0.65	0.66	0.91	0.93	0.91	0.93	0.96	0.97	0.94	0.94
	Recall	0.71	0.73	0.75	0.75	0.65	0.66	0.65	0.66	0.91	0.91	0.91	0.91	0.96	0.96	0.94	0.93
	F1-score	0.69	0.72	0.74	0.76	0.64	0.66	0.65	0.66	0.91	0.92	0.91	0.92	0.96	0.97	0.94	0.94
PLSDA	Accuracy	0.66	0.66	0.72	0.72	0.64	0.64	0.69	0.69	0.91	0.91	0.91	0.91	0.95	0.95	0.95	0.95
	Precision	0.64	0.64	0.72	0.74	0.61	0.61	0.67	0.68	0.91	0.91	0.91	0.91	0.95	0.95	0.95	0.95
	Recall	0.64	0.64	0.69	0.7	0.62	0.62	0.68	0.68	0.92	0.92	0.92	0.92	0.95	0.96	0.95	0.96
	F1-score	0.64	0.64	0.71	0.72	0.62	0.62	0.67	0.68	0.92	0.92	0.92	0.92	0.95	0.95	0.95	0.95

***Best model**

Table 4.11: Performance of classification algorithms with single and combination of two different preprocessing techniques of roasted coffee powder

Model	No of PCs		RAW		MC		MSC		SNV		SD		MC+SD		MSC+SD		SNV+SD	
			Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set	Training set	Test set
SVM	3	Accuracy	0.95	0.95	0.87	0.87	0.88	0.88	0.87	0.87	0.87	0.87	0.87	0.87	0.96	0.96	0.96	0.96
		Precision	0.95	0.94	0.88	0.88	0.88	0.87	0.87	0.87	0.89	0.88	0.89	0.88	0.96	0.94	0.96	0.94
		Recall	0.95	0.95	0.87	0.88	0.87	0.86	0.87	0.85	0.88	0.88	0.88	0.88	0.96	0.96	0.96	0.96
		F1-score	0.95	0.95	0.87	0.88	0.87	0.87	0.87	0.86	0.88	0.88	0.88	0.88	0.96	0.95	0.96	0.95
LDA	3	Accuracy	0.95	0.95	0.84	0.84	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.95	0.95	0.95	0.95
		Precision	0.95	0.96	0.84	0.85	0.87	0.87	0.87	0.87	0.88	0.88	0.88	0.88	0.95	0.95	0.95	0.95
		Recall	0.94	0.95	0.84	0.85	0.86	0.87	0.86	0.87	0.88	0.89	0.88	0.89	0.95	0.95	0.95	0.95
		F1-score	0.95	0.95	0.84	0.85	0.87	0.87	0.87	0.87	0.88	0.89	0.88	0.89	0.95	0.95	0.95	0.95
NN	3	Accuracy	0.91	0.91	0.83	0.83	0.85	0.85	0.86	0.86	0.83	0.83	0.87	0.87	***0.96**	**0.96**	0.95	0.95
		Precision	0.91	0.92	0.83	0.83	0.85	0.84	0.86	0.85	0.83	0.81	0.88	0.89	***0.96**	**0.97**	0.95	0.95
		Recall	0.9	0.9	0.82	0.84	0.85	0.85	0.86	0.86	0.84	0.84	0.88	0.88	***0.96**	**0.96**	0.95	0.95
		F1-score	0.9	0.91	0.82	0.84	0.85	0.84	0.86	0.86	0.84	0.82	0.88	0.88	***0.96**	**0.96**	0.95	0.95
RF	3	Accuracy	0.85	0.85	0.87	0.87	0.87	0.87	0.88	0.88	0.87	0.87	0.88	0.88	0.88	0.88	0.87	0.87
		Precision	0.85	0.84	0.86	0.86	0.87	0.86	0.88	0.87	0.88	0.89	0.88	0.88	0.88	0.87	0.88	0.89
		Recall	0.85	0.84	0.86	0.86	0.87	0.86	0.88	0.88	0.88	0.88	0.88	0.87	0.88	0.88	0.88	0.88
		F1-score	0.85	0.84	0.86	0.86	0.87	0.86	0.88	0.88	0.88	0.88	0.88	0.87	0.88	0.88	0.88	0.88
PLSDA	3	Accuracy	0.92	0.92	0.87	0.87	0.92	0.92	0.91	0.91	0.92	0.92	0.92	0.92	0.93	0.93	0.93	0.93
		Precision	0.92	0.92	0.86	0.87	0.92	0.92	0.91	0.91	0.92	0.92	0.92	0.92	0.94	0.93	0.94	0.93
		Recall	0.92	0.92	0.87	0.87	0.92	0.93	0.91	0.91	0.92	0.93	0.92	0.93	0.94	0.94	0.94	0.94
		F1-score	0.92	0.92	0.86	0.87	0.92	0.93	0.91	0.91	0.92	0.92	0.92	0.92	0.94	0.94	0.94	0.94

***Best model**

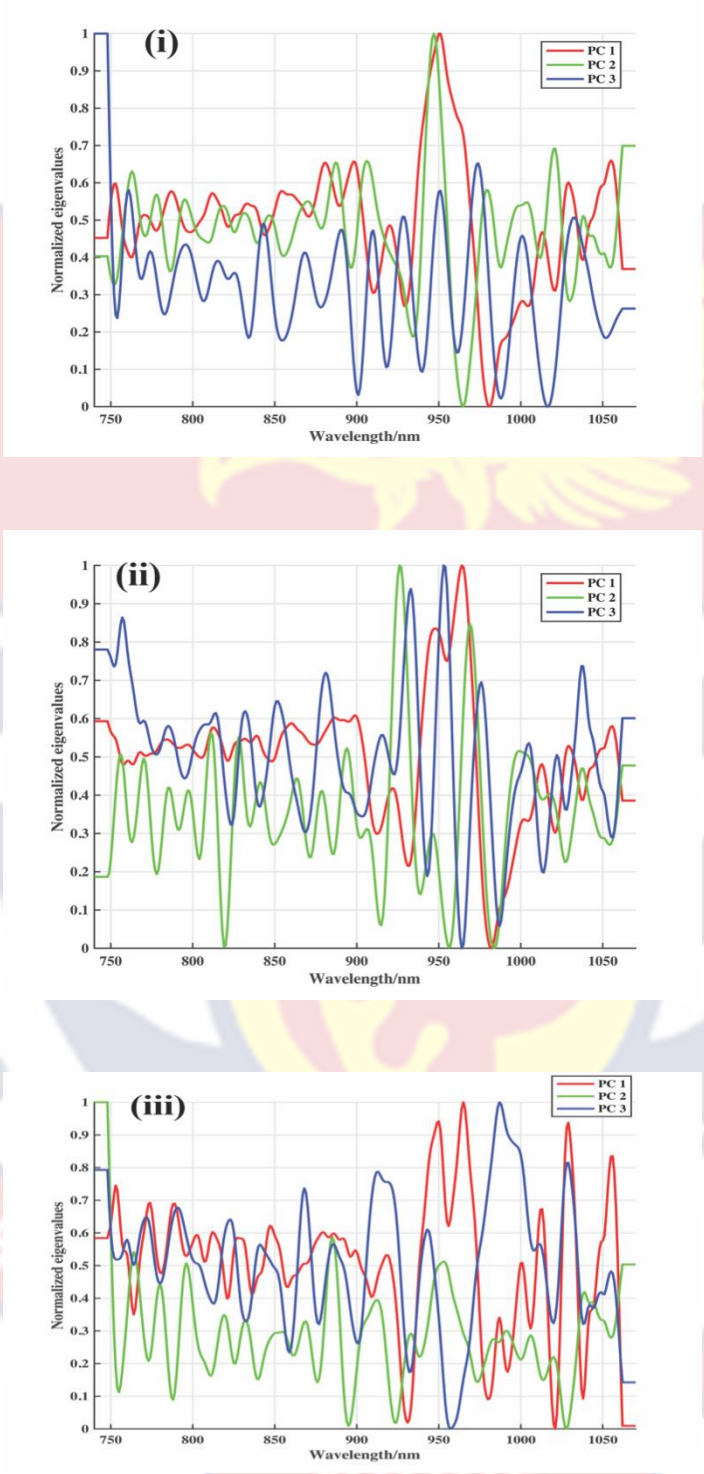


Figure 4.18: PCA loading score plot of geographical of (i) bean, (ii), roasted and (iii) roasted powder of coffee

4.4 General discussion

The spectra information of raw bean, roasted bean and powdered samples from five African countries displays a unique and similar profile. It is well recognized that chemical composition, the end of the visible spectrum, colour, and other characteristics have an impact on NIR spectra. As a result, the chemical makeup of samples from other countries varies, and this chemical composition is mostly influenced by the soil, weather, harvesting, and post-harvest processing. Countries that share common boundaries have closer and overlapping spectra. These observations could be a result of similar farming methods, pre-and post-harvest activities that influence coffee bean quality attributes. Bean, roasted and roasted powdered coffee bring about differences in their proximate composition which results in spectra variations in these categories. Some chemical parameters like moisture content and protein were lost when heat was applied to them. However, the rapid classification of coffee beans from different countries was possible irrespective of the state (bean, roasted and powdered samples).

Five preprocessing methods were employed in this study; FD, SD, MC, MSC, and SNV were used and MC provided the best treatment results for PCA with 99.76% and 99.88% for bean and roasted coffee powder respectively, while MC and MSC provided the best PCA score of 99.78% for roasted coffee. The score plot of the PLS-DA model built from spectra for the coffee types of five African countries showed a separation with the following results; SD-PLSDA had the best accuracy and F1-score of (0.87, 0.87) for beans and (0.91, 0.91) for roasted coffee while MSC-PLSDA had (0.92, 0.93) for roasted coffee powder. The PCA loading

plot as shown in Figure 4.18 reveals the major wavelength denoted by peaks that influenced the accurate classifications of the coffee samples. These wavelengths reveal very important unique chemical structure that differentiates the various groups of coffee bean samples. As seen in Figure 4.18, the major peaks for raw, roasted and powder coffee beans were centered around 742 – 790, 830 – 890 and 920 – 960 nm which corresponds to the third overtone region which are known form of CH_3, CH_2, CH, and some form of RNHR and RNH_2 (Barbin et al., 2014). These wavelengths are associated with aromatic functional groups in coffee and especially the third overtone of CH with its associated OH stretch of H_2O are normally assigned to phenols and antioxidants in coffee (Kljusurić et al., 2016).

For the classification models, SVM, NN, RF, and PLS-DA models were developed with a single or combination of two different spectra preprocess techniques for bean, roasted and powder coffee. The second derivative with PLSDA and MC plus the second derivative with PLS-DA outperformed the other classification models, achieving an accuracy of 0.87 and an FI-score of 0.88 for the coffee bean. With an accuracy of 0.97 and an F1-score of 0.97, SNV plus the second derivative had the best SVM model for the roasted coffee samples. For roasted coffee powder, MSC + second derivative had the best NN model with an F1-score and accuracy of 0.96. These models were regarded as the finest, however, SNV+SD-SVM exceeded them all. It could be that SD only picks up minor spectra changes, but SNV corrects multiplicative interferences of light and particle size. With a restricted number of training samples, SVM has also demonstrated good performance in classifying high-dimensional data.

4.5 Conclusion

Geographical differentiation of African Robusta coffee types (bean, roasted, powdered) could be accomplished using handheld near-infrared spectroscopy in combination with an appropriate multivariant classification algorithm in chemometrics. Accuracy and the F1-score were used to evaluate the model's effectiveness. The best classification models were developed for raw bean coffee, SD-PLSDA, and MC+SD-PLSDA, and they had an F1-score of 0.88 and an accuracy of 0.87 respectively. SNV+SD-SVM had 0.97 accuracy and 0.97 F1-scores for roasted coffee beans, while MSC+SD-NN had 0.97 accuracy and 0.97 F1-scores for roasted coffee powder. According to the results, it is possible to geographically differentiate between various types of African coffee using analytical data from handheld NIR spectroscopy. This technique could benefit coffee bean producers in developing countries like; Ghana, Uganda, Burkina Faso, Ivory Coast and processors mostly in industrialized countries like Japan, Holland, and Germany.

CHAPTER FIVE

PORTABLE NIR SPECTROSCOPIC APPLICATION FOR COFFEE INTEGRITY AND DETECTION OF ADULTERATION WITH COFFEE HUSK

Vida Gyimah Boadu[1,2], *Ernest Teye[1] Charles Amuah[3,] Francis Padi Lamptey[1,4] and Livingstone Kobina Sam-Amoah[1]

[1]University of Cape Coast, College of Agriculture and Natural Sciences, School of Agriculture, Department of Agricultural Engineering, Cape Coast, Ghana

[2]Akenten Appiah Menka University of Skill Training and Entrepreneurial Development, Department of Hospitality and Tourism Education, Kumasi Ghana,

[3]University of Cape Coast, College of Agriculture and Natural Sciences, School of Physical Sciences, Department of Physics, Cape Coast, Ghana

[4]Cape Coast Technical University, Department of Food Science and Postharvest Technology, Cape Coast, Ghana

*Corresponding author

Published with slight modification, Journal of Processes (Scopus/Sci, IF = 3.52)

Statement of contribution of Joint Authorship

Boadu, V. G.: **(Candidate)**

Conceptualised the topic, established methodology, data collection and analysis, preparation of tables and figures, writing and compilation of the original manuscript.

Teye, E.: **(Principal Supervisor)**

Conceptualised the topic, established methodology, supervised and edited the manuscript and co-author of manuscript.

Sam-Amoah, L. K.: **(Co-Supervisor)**

Conceptualised the topic, supervised and edited the manuscript and co-author of manuscript.

Charles L. Y Amuah: **(Research Team member)**

Validation, visualization and data curation and co-author of manuscript

Francis Padi Lamptey: **(Research Team member)**

Validation, visualization and data curation and co-author of manuscript

Abstract

Reliable and user-friendly discrimination of coffee bean integrity and quantification of adulterate would be vital for ensuring consumer trust in quality control and traceability management. In this research, short-wave NIR spectroscopy coupled with chemometric data analysis was employed. Different pre-processing treatments (multiplicative scatter correction, MSC; standard normal variant, SNV; first derivative, FD) together with multivariate techniques; support vector machine (SVM), linear discriminant analysis (LDA), neural network (NN), and random forest (RF) were comparatively assessed using accuracy and correlation co-efficient (R) for discrimination accuracy and quantification efficiency respectively of the models. The results showed that FD-LDA model had 97.78 % and 100 % in both calibration set and prediction set. While, SPA-PLS model had R = 0.9711 and 0.9897 in both calibration set and prediction set. The outcome of this study showed that Short-wave NIR spectroscopy could be used for coffee integrity examination.

Keywords: Coffee, Adulteration, Pure, Short-wave NIR spectroscopy, Chemometric

5.1 Introduction

Food fraud has now been reported in every food commodity and has become a universal food safety concern with adulteration toping all forms of food fraud (Ting et al., 2020). Food adulteration form of food fraud is the action of manufactures in increasing the profit of authentic product with low quality of products by removing expensive components to increase the amount of the product with cheap materials (Spink & Moyer, 2011). Often this fraudulent act is targeted to widely consumer commodities like coffee. Due to its ease of adulteration, which includes techniques like adding flavours or aromas and using an unknown addictive to boost its volume, among others, coffee, unfortunately, has one of the highest reported numbers of fraud instances among beverages (Kamiloglu, 2019; Thorburn Burns et al., 2017). Roasted coffee is frequently and in various ways adulterated. In order to reduce the price of coffee blends, it may be necessary to alter the quality of the beans (taking into account the species, region of origin, and defective beans) as well as to add additional ingredients such as coffee husk, stems, maize, chicory, barley, wheat middling, brown sugar, soybeans, rye, and triticale (Toci et al., 2016).

Coffee is an important raw material traded globally and one of the most popularly consumed beverage (Choi et al., 2016). Food safety concerns has recently been on a high agenda of many consumers and coffee quality assessment is impelled by the need to supply the consumers with a consistently high-quality product at an affordable price (Barbin et al., 2014). This therefore has called for a

far greater scrutiny to provide undeniable coffee integrity. However, some form of adulteration goes on unnoticed with majority being adulteration of the roasted coffee with coffee husk (Prodolliet et al., 1995). This act is further influenced by the increasing demand of coffee propelled by its consumption in emerging markets around the world especially Russia, Australia and South Korea (Catelani et al., 2018). Since only the person handling the food is aware that the item has been altered, they are the only ones with knowledge, but they might not have the expertise to assess whether such alteration puts the customer at risk. These practices are prohibited everywhere and not only have negative economic effects (Núñez, Saurina, et al., 2021).

Therefore, coffee roasters, consumers and quality control officials need to rise up to the task of ensuring the integrity of coffee consumed world-wide and not only focused on massive production to catch-up with the short fall in supply. Hence, user-friendly and suitable analytical techniques are needed, for quality, safety and economic reasons as well as to prevent coffee fraud and ensure integrity, thus promoting consumer confidence and encouraging and maintaining coffee beverage consumption world-wide. These were made possible by employing different wet chemistry methods including compositional data analysis (volatile compounds, fatty acid profile, chlorogenic and caffeine content) (Esteban-Díez et al., 2007), analysis of total xylose (Prodolliet et al., 1995) and other known composition. Furthermore, other researchers have used methods such as chromatographic and, enzymatic methods while anion-exchange chromatography with pulsed amperometry detection is mostly preferred as the most powerful technique

(Prodolliet et al., 1995). However, these aforementioned techniques are quite expensive, elaborate, time-consuming and often not applicable for onsite real time analysis.

Researchers have used different analytical methods to detect adulteration of coffee. UV-Vis spectroscopy was employed by (de Carvalho Couto et al., 2021) to successfully detect adulteration in the form of husk and sticks in the ground roasted coffee samples. Again the following equipment were used by other researchers in adulteration detection in coffee; normal-phase HPLC with florescence (Tavares et al., 2016), PCR-capillary electrophoresis (Uncu & Uncu, 2018) and Gas chromatography solid phase micro-extraction (Oliveira et al., 2009).

NIR technology offers a great replacement for the conventional methods used for industrial coffee quality control. Due to sensitivity, non-destructiveness, speed, minimal sample preparation requirements, and lack of the use of toxic solvents, NIR data reveals itself as a potentially efficient tool for the detection of coffee adulteration. Various food items have been found to be adulterated using this technique. (Rodriguez-Saona & Allendorf, 2011). NIR was utilized to identify coffee that contained adulterated arabica beans made of barley(Ebrahimi-Najafabadi et al., 2012), corn (Winkler-Moser et al., 2015), and Robusta (Correia et al., 2018; Pizarro, Esteban-Díez, & González-Sáiz, 2007). The non-destructive detection of coffee containing adulterants such as barley and maize was done using FT-NIR spectroscopy (Chakravartula et al., 2022; Ebrahimi-Najafabadi et al., 2012). Portable NIR was used by (Ebrahimi-Najafabadi et al., 2012) to identify 2% w/w of barley while (Winkler-Moser et al., 2015) detected 5% of corn in roasted

coffee. (Correia et al., 2018) also demonstrated it effectiveness to quantify 5-8wt% of corn, sticks and Robusta coffee in arabica coffee.

Even though it is speedier than wet chemistry, it can be difficult, particularly in developing nations, to rely only on pricey, stationary laboratory-based equipment. For these reasons, the coffee authenticity drive is looking for faster, low cost and onsite method and portable NIR spectroscopy that could be a very useful. The development in NIR downsizing has opened up new possibilities for NIR applications ideal for in-person, lab, and industrial examination. Large, stationary laboratory-based NIR tools have become portable as a result of this. Portable NIR is more affordable, less complicated to use, and requires fewer equipment, highly resistant to mechanical stress than convectional, lab-based spectrum measuring spectrometers. These instruments' other benefits include their movability, strength, and ability to permit in-field and product-to-product evaluation and regulation. Some of the additional advantages over the traditional instrument designs include on line and in situ analyses. They also offer significant advantages in terms of size, weight, robustness, spectral range and low manufacturing process (Kademi et al., 2019). Again, they are useful in situations that require an emergency response and process monitoring applications. However, their disadvantages include higher detection limits, lower sensitivity, a strong influence of environmental factors on instrument performance and high possibility of a sample contamination in the field (Gałuszka et al., 2015).

Presently, several techniques utilizing portable spectroscopy have been developed for quick and in-field food analysis. Portable NIR and Raman

instruments are receiving particular interest because of their analytical performance, affordability and capacity for real- time field analysis (Dégardin et al., 2017). The two devices record information on the characteristics on the physical and chemical properties of the sample combined with irrelevant information and interfering signals. NIR is based on the absorption of light whereas Raman spectroscopy is based on the scattering phenomenon. The advantage of portable NIR over Raman spectroscopy is that, the latter has drawbacks including the adverse effects of fluorescence, lack of sensitivity and less useful for specific product identification (Ciza et al., 2019).

Santos et al. (2021) employed a portable NIR spectroscopy to discriminate crude oils and derivatives and quantify them in mixtures with used motor oils, gasoline and diesel. It was also utilized by (Zhao et al., 2021) to quantify soil plastics pollution levels and by (Kranenburg et al., 2022) to identify forensic materials. In the field of coffee research, (Mutz et al., 2023), developed a supervised classification model for differentiating between Robusta and arabica coffee using portable NIR spectroscopy and chemometrics techniques. Correia et al. (2020) also proposed a new analytical methodology to observe the qualities of Robusta coffee cultivated in agroforestry systems using portable micro NIR spectroscopy in tandem with sensory analysis.

It has not been utilized frequently, as far as we know, for onsite identification of adulterants in coffee. Consequently, the goal of the study was to assess the effectiveness of using chemometrics and portable NIR spectroscopy together to detect coffee husk in samples of Robusta coffee.

5.2 Materials and Methods

5.2.1 Sample collection

Robusta coffee and coffee husk were gathered for this study from several locations in African countries that produce coffee. These samples consist of 90 adulterated samples created by accurately adding coffee husk (5-30% w/w) to Robusta coffee, 40 samples of coffee husk, and 40 samples of Robusta coffee. According to (Vasconcelos et al., 2007), the coffee samples were roasted for 1 hour at 200^{o}C. To make the coffee husk's colour resemble that of roasted coffee, it was first roasted. A home electric coffee bean roaster was used to roast the beans and husk (Jiawanshun, China). Using a multipurpose grinder, the coffee husk and Robusta coffee were ground separately (QE- 100, Zhejiang YiLi Co, Ltd, China). All of the sample groups were well-labelled and transported to the school of Agriculture laboratory at the University of Cape Coast.

5.2.2 Spectra collection

In the laboratory, a portable NIR spectrometer (SCIO^{TM}) in the range of 740 -1070nm was used to capture the spectrum of each sample from the three different coffee categories, (Robusta coffee, husk coffee, and adulterated coffee, all in the powdered state) with a resolution of 1 nm for data collection assisted by a smartphone (Samsung A21). About 50 g of each sample was gathered in a ziplock bag and subjected to five scans and the spectra were averaged to provide a mean spectrum as the original spectrum of sample used. The scanning was carried out in a steady state of humidity and an ambient temperature of 31°C at the school of Agriculture laboratory, university of Cape Coast.

5.2.3 Chemical composition determination

The proximate composition, antioxidants and polyphenols of all the samples were carried out using standard methods according to the standard used by other authors (AOAC, 1995). All the parameters of the samples were carried out in triplicate and average to represent one sample.

5.2.4 Spectra data processing

All the computations, chemometric analysis and graphics were done with MAT LAB (2021a, MathWorks Inc., USA) using the windows 10 Basic software package for all data processing. In this study, three pre-processing techniques—first derivative (FD), multiplicative scattered correction (MSC), and SNV (standard normal variate) were utilized in comparison. Other authors have employed these pre-processing techniques. Consult the earlier research of other writers for more information on the theory underlying the pre-processing techniques that were employed (Bi et al., 2016; Miao et al., 2021).

5.2.5 Principal component analysis (PCA)

After pre-processing, PCA was employed as an unsupervised pattern technique of identification to analyse potential information trends in a dimensional space in the form of a score plot. By compressing the data into a major component that includes practical interpretable variables, it is a well-known approach for lowering the dimension of the data matrix. The top three PCs in a PCA hold crucial information and frequently initiate crucial information with minimal to no redundant data.

5.2.6 Data partitioning

The spectral data set for pure Robusta coffee (40), adulterated samples (90) and pure coffee husk (40) was downloaded individually, and two sets (the training set and the test set) of each category were created. A total of 58 samples were chosen as the testing set when the model was being tested, and 112 samples were chosen as the training set when the model was being built. Three spectra were randomly chosen from every five samples as a training set, while the other samples were utilized as a test set to prevent bias in the division.

5.2.7 Multivariate data modelling

The study's main objective is to identify and quantify coffee promptly. Accurately recognizing Robusta coffee, coffee husk, and adulterated coffee is the identification problem. Support vector machine (SVM), linear discrimination analysis (LDA), neural network (NN), and random forest (RF) were some of the identification techniques that were compared in order to solve this problem. Please check the authors for more details on the ideas underlying these identification techniques (Teye et al., 2013; Tian et al., 2022; Wilamowski, 2009).

To compare various forms of partial least square regression (PLS, iPLS, biPLS, GaPLS, and SPA-PLS), the quantity of coffee husk adulteration in Robusta coffee was determined. These different multivariate modelling approaches were compared and evaluated after their application using the test set performance data because each has strengths and shortcomings of its own. Please refer to the authors for further information on the ideas behind the quantitative models employed in the study (Inácio et al., 2013; Miao et al., 2021).

5.2.8 Model development evaluation

In order to evaluate how well the qualitative and quantitative models performed, various statistical approaches were applied. As done by previous writers, the true positive rate (TPR), true negative rate (TNR), false positive rate (FPR), and false negative rate (FNR) were used to examine the performance of the identification models. (Chen et al., 2019). The predictive performance of the quantitative models was examined using the correlation coefficients of calibration and prediction sets (R), as well as the root mean square errors of cross-validation and prediction (RMSECV and RMSEP), which were employed by other researchers. (Kamruzzaman et al., 2013; Teye & Amuah, 2022).

5.3 Results and Discussion

5.3.1 Wet chemistry results

There was a significant difference in all the parameters measured for coffee and coffee husk except for moisture. Ash (6.26%), lipids (6.23%) and Fibre (6.19%) in coffee husk were higher than coffee. Proteins, carbohydrates, polyphenols and antioxidants were higher in the coffee than the coffee husk. Comparing proximate composition of the husk to literature, ash was higher than values recorded by (5.4%) and in range with (6.2%) (Bekalo & Reinhardt, 2010). Protein was higher than all values recorded in literature (7.0% and 11.0%) for (Gouvea et al., 2009) and (Franca & Oliveira, 2009) respectively. Addition of coffee husk to the coffee decreases the amount of antioxidants and polyphenols since coffee has higher amount of polyphenols and antioxidant as recorded in Table 5.12.

Table 5.12: Chemical composition of Robusta coffee and husk

Parameters	Robusta coffee	Husk
Moisture (%)	2.54 ± 0.32 [a]	3.24 ± 0.59 [a]
Ash (%)	3.62 ± 0.28 [a]	6.26 ± 0.20 [b]
Protein (%)	16.19 ± 0.11 [a]	14.59 ± 0.19 [b]
Lipid (%)	5.46 ± 0.29 [a]	6.23 ± 0.10 [b]
Fibre (%)	5.76 ± 0.09 [a]	6.19 ± 0.06 [b]
Carbohydrate (%)	68.98 ± 0.66 [a]	66.73 ± 0.38 [b]
Polyphenols (mg/kg)	4373.30 ± 65.80 [a]	354.30 ± 29.0 [b]
Antioxidant (mg/kg)	4416.80 ± 18.50 [a]	739.67 ± 7.43 [b]

Values represent mean ± SD of at least three replicates. Means that do not share letters in same rows are significantly different at $p < 0.05$.

Antioxidants and polyphenols may significantly improve quality of life by assisting in the prevention or delaying of degenerative diseases (Svilaas et al., 2004; Williamson, 2017). Husk and other impurities lower the quality of ground and roasted coffee, which adversely affects the beverage's flavor, fragrance, acidity, bitterness, and other sensory qualities (Meilgaard et al., 1999; Tavares et al., 2012).

5.3.2 Spectral profile examination

As depicted in Figure 5.19, the spectral profile of the coffee samples used in this research had a distinct fingerprint. The raw spectra for Robusta coffee, coffee husk, and adulterated coffee are displayed in Figure 5.19(a). This chart made it clear that Robusta coffee's spectrum was distinct from that of coffee husk and adulterated coffee. Figure 5.19(b) showed that all three samples (Robusta coffee, coffee husk, and adulterated) overlapped at 790 and 980nm. This indicates that the samples exhibit similar physical and chemical properties at these wavelengths, and Table 5.12 illustrates this phenomenon. Coffee is frequently adulterated with coffee husk because of these similarities in proteins, carbohydrates, lipids and caffiene.

Moreover the wavelength range from 740 to 900nm represent 3rd overtone Aromatic C-H, C-C and C=C which is associated with proteins, carbohydrates and lipids. Also in the region,740 to 1070 nm 2nd and 3rd overtone RNH and ROH represent proteins, antioxidants and polyphenols of coffee samples used (Barbin et al., 2014). The entire utilized wavelength range (740–1070 nm) contains aromatic C–H, C–C, C=C, N–H, and O–H chemical bonds that may be connected to polyphenols, antioxidants, proteins, lipids, and carbohydrates. To ensure that the samples were distinct from one another, the mean spectrum of the scanned samples was done as shown in Figure 5.20. It was seen that the Robusta coffee samples clearly distinguished themselves from the coffee husk and adulterated samples at wavelengths between 750nm and 1050 nm in Figure 5.20a. When light of various wavelength are radiated to organic matter, a portion of the light at certain wavelength are absorbed. The amount of light absorbed depends on the composition of the irradiated organic materials in the case of coffee, husk and adulterated samples. They have different composition such as polyphenols, proteins, lipids, carbohydrates and moisture. This causes different absorption responses for each. This band's range corresponded to the band previously observed in spectra for roasted coffee and coffee husk (Reis et al., 2013). In Figure 5.20b, aborbance was lower for coffee and higher for coffee husk. The adulterated samples were found in between coffee and husk that could be variatons in their chemical composition.

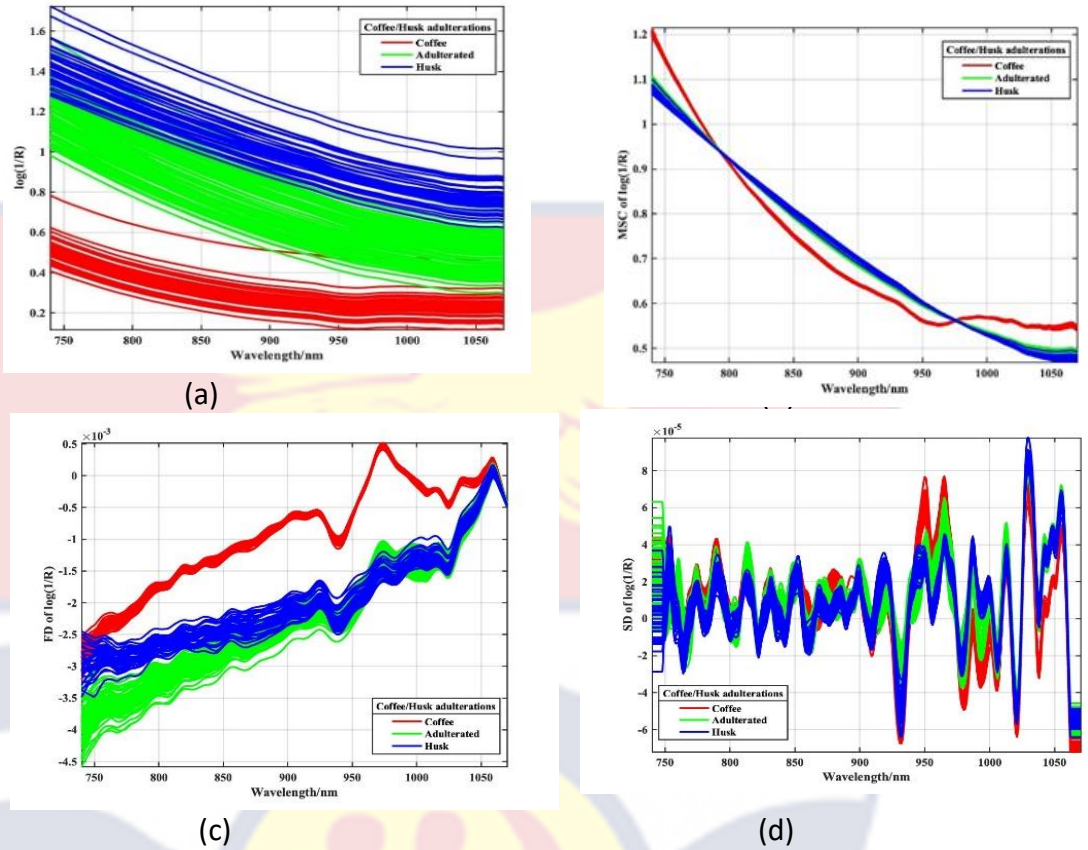


Figure 5.19: Raw spectral profile of Robusta coffee and adulterants: (a) raw, (b) MSC, (c) FD and (d) SD

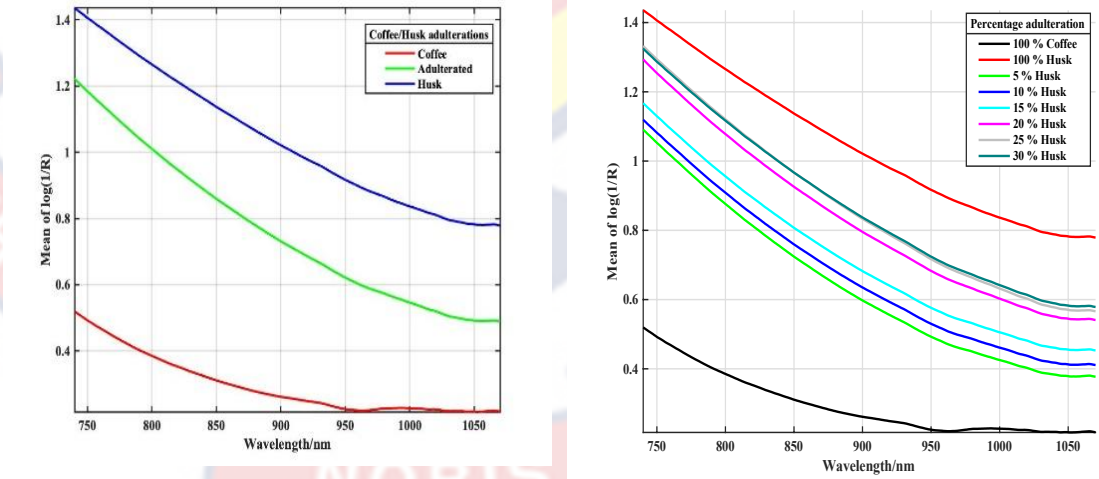


Figure 5.20: Mean coffee category spectral profiles in their (a) raw form and (b) adulterated levels

5.3.3 Principal component analysis (PCA)

In this study, roasted Robusta coffee and coffee husk were subjected to principal component analysis to reveal the underlying natural pattern as a PCA score plot. Figure 5.21 displays the PCA for samples of roasted coffee, coffee husk, and adulterated coffee. The PCA is an unsupervised pattern. It typically determines the primary phenomena in the dataset and the direction of variation that is most important in the data set space (Sun et al., 2017).

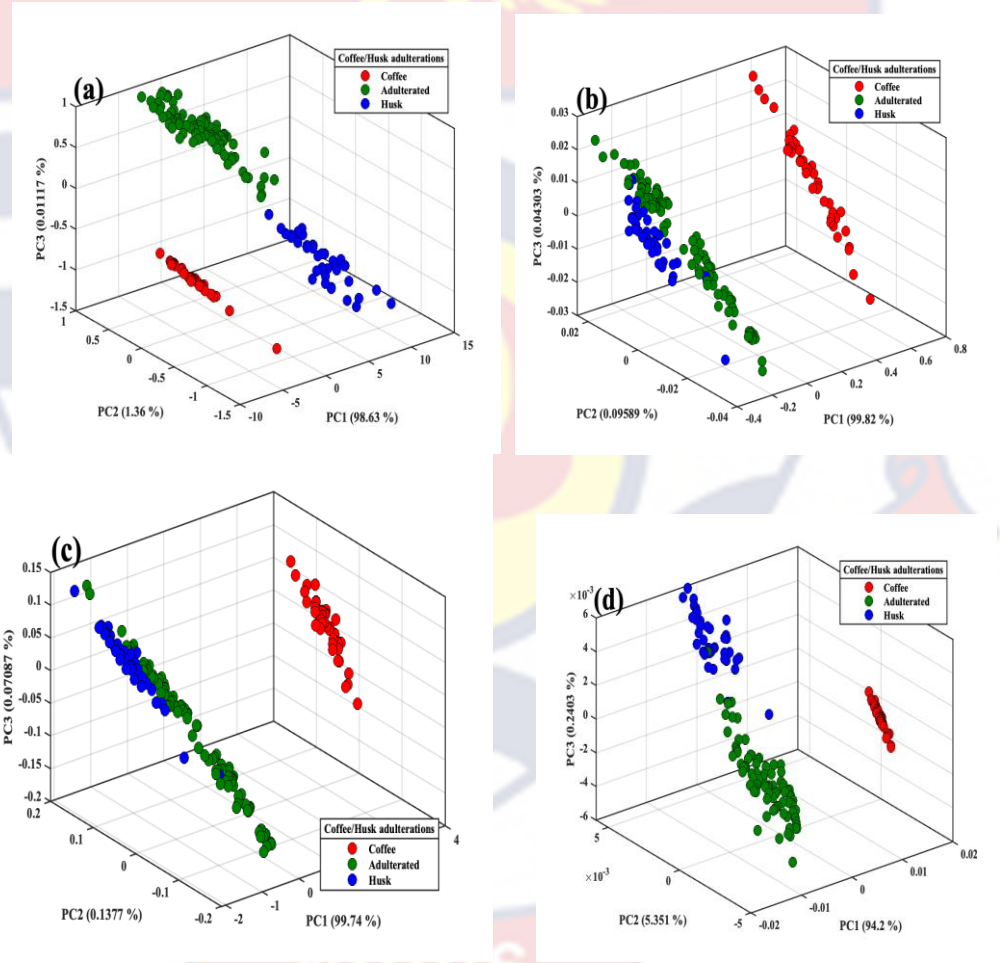


Figure 5.21: PCA score scatter for Robusta coffee and adulterants: (a) raw, (b) MSC, (c) SNV and (d) FD

With pretreatments result of 99.96%, MSC outperformed all other preprocessing methods, followed by FD (99.79%) and SNV (99.95%). As can be

observed in the PCA score plot in Figure 5.21, all three samples were clearly clustered for MSC as shown in Figure 5.21 (a). MSC outperformed alternative pretreatments because it can correct the scatter, addictive and multiplicative effects than other pretratment (Dhanoa et al., 1994).

The PCA eigenvector plot explained the cluster trend observed in the PCA score plot. As can be seen from Figure 5.22. The key peaks responsible for the neat clustering are situated between 950 and 1000 nm for PC1, 920 and 930 nm for PC2, 930 and 960 nm, 960 and 980 nm, and 990 and 1100 nm for PC3, according to the loading plot of FD-PCA Eigenvectors. The second and third overtone regions' RNH, ROH, CH2, and CH3 are represented by these significant peaks (Barbin et al., 2014).

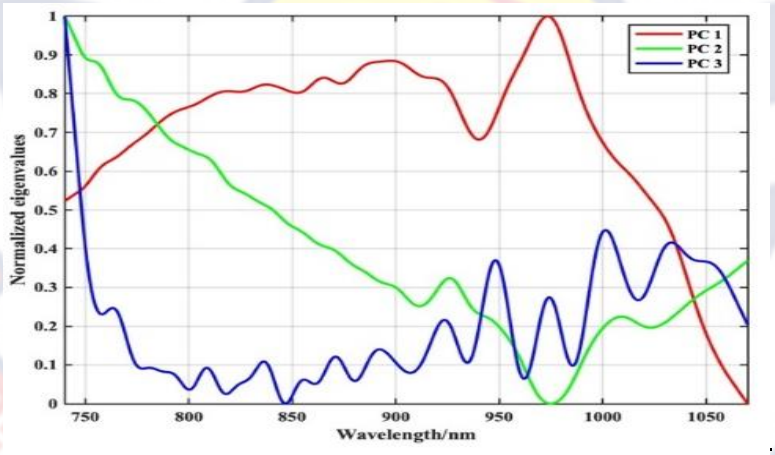


Figure 5.22: FD-PCA Eigenvectors for the coffee samples

5.3.4 Identification model

Identification models were also constructed and assessed in comparison. Support vector machine (SVM), linear discrimination analysis (LDA), neural network (NN), and random forest (RF) were among these models. The models produced the best results with preprocessing techniques like FD, MSC, and SNV.

Figure 5.23 shows that the performance of every identification model was considerably above a 90% identification rate. The first derivative spectra preprocessing treatment, however, outperformed the others when the raw spectra data set was processed, achieving 97.78% and 100% in both the calibration set and the prediction set using LDA, as shown in Table 5.13.

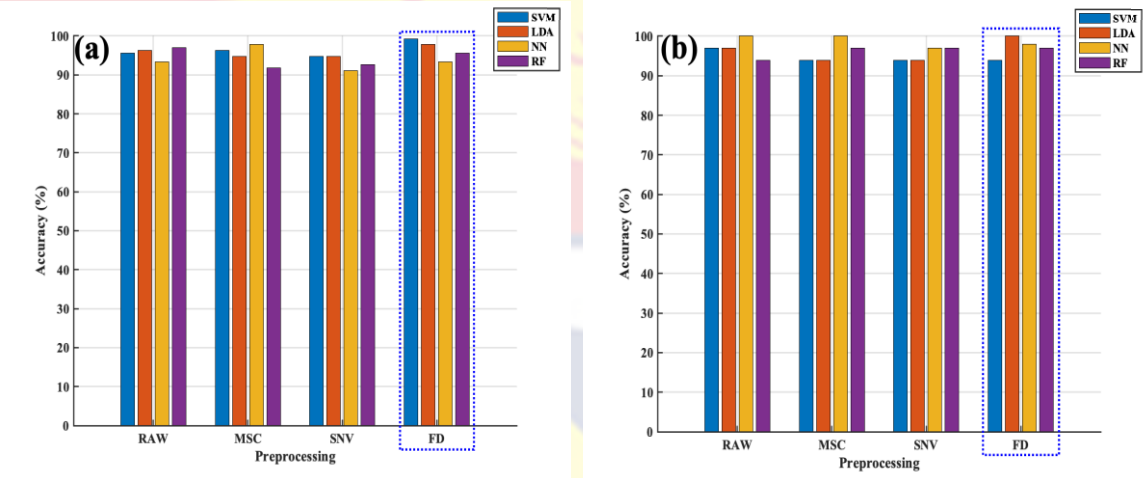


Figure 5.23: Accuracy in (a) calibration set and (b) prediction set for SVM, LDA, NN and RF models of Coffee samples

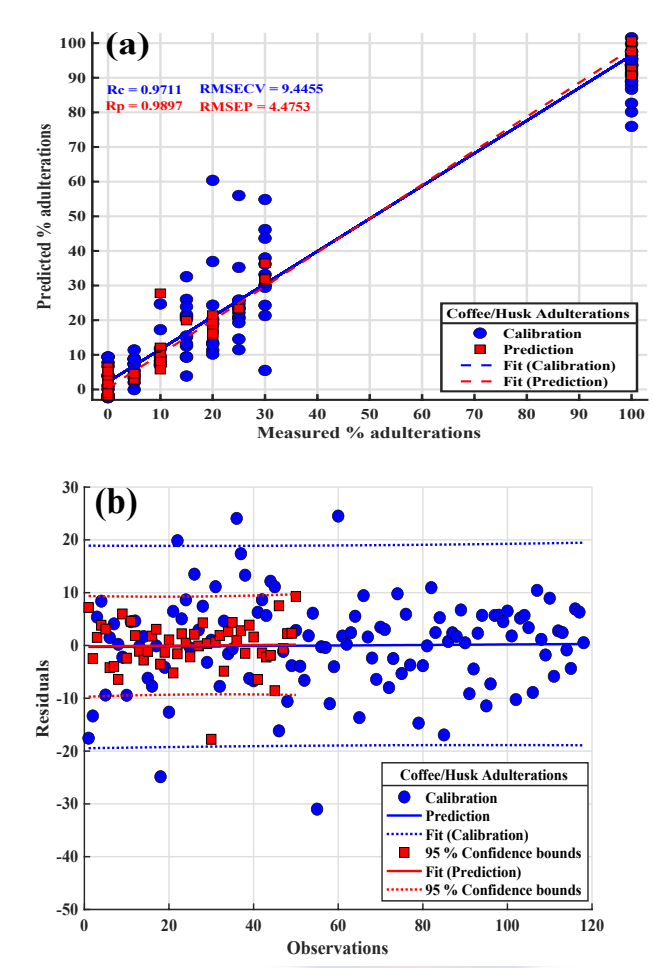
Table 5.13: Optimum discrimination accuracy for the models developed

Model	Preprocessing		Accuracy (%)	Error (%)	TPR (%)	TNR (%)	FPR (%)	FNR (%)
SVM	FD	Calibration	99.26	0.74	100.00	100.00	0.00	0.00
		Prediction	93.94	6.06	100.00	100.00	0.00	0.00
LDA	FD	Calibration	97.78	2.22	100.00	100.00	0.00	0.00
		Prediction	100.00	0.00	100.00	100.00	0.00	0.00
NN	FD	Calibration	96.30	3.70	100.00	99.05	0.00	0.95
		Prediction	96.97	3.03	100.00	100.00	0.00	0.00
RF	FD	Calibration	95.56	4.44	100.00	100.00	0.00	0.00
		Prediction	96.97	3.03	100.00	100.00	0.00	0.00

Again, the model was assessed by the true positive rate (TPR), true negative rate (TNR), false positive rate (FPR) and false negative rate (FNR). It is always necessary for the false positive rate to be decreased while the true positive rate is

maximized (Simmons et al., 2008). The true positive rate was 100% for all the models with less than 1% for the false negative rate. The capacity of linearly discriminating functions, which clearly highlights the ratio of class variance and lowers the ratio of within-class variation, is what gives LDA its best performance (Q. Chen et al., 2011).

5.3.5 Multivariate quantification of adulterant



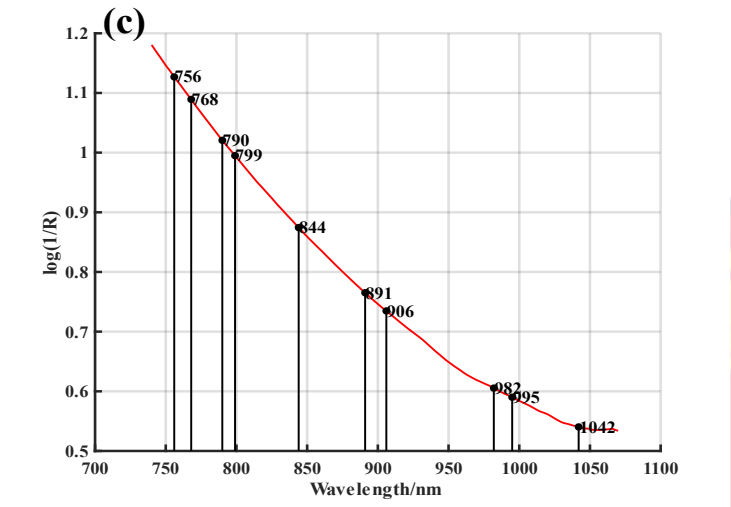


Figure 5.24: Scatter plots of SPA-PLS model with NIR estimation and experimental concentration (a) and residual plot (b) and (c) model plot

To more precisely quantify the authenticity and levels of adulteration in the samples, various partial least square regression models, including PLS, iPLS, biPLS, Si-PLS, GaPLS, and SPA-PLS, were utilized for quantitative analysis. PLS could provide accurate predictions and extract relevant information. These methods were improved and contrasted. With R=0.97 for the calibration set, R=0.98 for the prediction set, and RPD (ratio performance deviations) of 7.05, the results in Table 5.14 showed that the SPA-PLS model provided the best adulteration prediction results for the samples that had been tampered with. These results concur with those of other researchers who tackled a related problem using even a wider wavelength (de Carvalho Couto et al., 2021; Reis et al., 2013). Among these models, the SPA (successive projections algorithm) chooses variables with the least amount of collinearity by using straightforward projection operations. It is a cutting-edge variable selection algorithm that has also been used to successfully address collinearity issues (Li et al., 2014).

The optimum spectrum selection of the essential wavelengths that produce the research's accurate results, which are 756-768 nm, 790-799nm, 844nm, 891-906nm, 982-995nm, and 1042nm, is illustrated in Figure 5.24 (c) to explain this phenomenon. These wavelengths match the various chemical characteristics of the coffee that set it apart from the coffee husk. More significantly, the third overtone CH_3, CH_2, and ROH ranges from 756-768nm, 790-799nm, 844,891-906nm, and 844,891-1042nm are linked to carbohydrates, while the second overtone R-NH, aromatic-CH, CH_2, and CH_3 ranging from 982-995nm and 1042nm are linked to proteins, lipids, polyphenols, and antioxidants. The aforementioned qualities offer incredibly distinctive chemical traits that can be utilized to precisely identify the grade classes and make predictions about the integrity of the coffee. This provides more evidence for the claim made by Correia et al. (2018) that polyphenolic chemicals were crucial in the identification and application of their spectrum data, which allowed for the precise observation and determination of distinct samples.

Table 5.14: Comparison of different PLS quantification models

Models	Variables	Calibration set			Prediction set			RPD
		R	RMSECV	Bias	R	RMSEP	Bias	
PLS	331	0.9626	10.7187	0.9867	0.9858	5.2485	0.7422	6.0179
iPLS	16	0.9701	9.6209	0.2049	0.9704	8.2453	-0.8647	3.8306
biPLS	49	0.9682	9.9263	0.1241	0.9708	8.2086	-0.6206	3.8478
Si-PLS	50	0.9799	7.888	0.0056	0.9824	6.0610	-1.4073	5.2111
GaPLS	331	0.9674	10.0189	-0.1527	0.9792	7.1855	-3.1656	4.3956
SPA-PLS	10	0.9711	9.4455	0.8695	0.9897	4.4753	0.6329	7.0576

The ratio of random variation in the samples to the degree of expected prediction errors is described as residual prediction deviation (RPD). RPD is more advantageous when comparing models on different data sets or in absolute terms.

It was calculated for the calibration and prediction sets and produced values that showed how well or poorly the models were able to predict outcomes. Values of residual prediction deviation greater than 2.4 indicate calibration models with good predictive capacity, whereas values less than 1.5 are considered poor (Botelho et al., 2013; Porep et al., 2015). The SPA-PLS model, which has the best PRD value of the other models at 7.05 as shown in Table 5.14, has all of the recorded values higher than 2.4. Figure 5.24b shows the residual plot for the SPA-PLS for the both the calibration and prediction plots with 95% confidence bounds. This plot shows the differences between the measured percentage of adulteration and the predicted percentage of adulteration. More than 90% of the calibration data scatters around the mean (fit) line with the remaining outside the confidence bounds. These data points can be considered to be outliners. The same can be said for the prediction data.

5.4 Conclusion

This research work has demonstrated the potential of miniaturized short-wave NIR spectroscopy for coffee integrity in Ghana and could be used as a technique for rapid, onsite and affordable coffee examination. The best technique was found to be first derivative (FD) pre-processing modelled with linear discriminate analysis (FD-LDA). Thus; FD-LDA was superior (97.78 % and 100 % in both calibration set and prediction set) to the others used for qualitative determination of coffee integrity (discrimination of pure coffee from adulterated ones). While for quantitative detection of coffee adulterant (quantification of the percentage of adulterant; 5 % - 30 %) in authentic coffee, SPA-PLS model had R =

0.9711 and 0.9897 in both calibration set and prediction set. The results obtained in this study showed only the feasibility and additional research are required to confirm the performance across different locations and varieties.

CHAPTER SIX

MICROBIAL AND SAFETY PARAMETERS OF COMMERCIALLY SOLD COFFEE IN GHANA

*Vida Gyimah Boadu[1,2], *Ernest Teye[1] & L. K. Sam-Amoah[1]

[1]University of Cape Coast, College of Agriculture and Natural Sciences, School of Agriculture, Department of Agricultural Engineering, Cape Coast, Ghana

[2]Akenten Appiah Menka University of Skills Training and Entrepreneurial Development, Department of Hospitality and Tourism Education, Kumasi Ghana,

*Corresponding authors

Under review: African Journal of Food science

Statement of contribution of Joint Authorship

Boadu, V. G.: **(Candidate)**

Conceptualised the topic, established methodology, data collection and analysis, preparation of tables and figures, writing and compilation of the original manuscript.

Teye, E.: **(Principal Supervisor)**

Conceptualised the topic, established methodology, supervised and edited the manuscript and co-author of manuscript.

Sam-Amoah, L. K.: **(Co-Supervisor)**

Conceptualised the topic, supervised and edited the manuscript and co-author of manuscript.

Abstract

Coffee is the most commercial crop in the economy of many countries in the world. Coffee contaminants include the most common toxigenic fungus taxa (*Aspergillus)*,

according to studies on the microbiology of coffee powder. During the processes of cultivation, harvest, transportation, and storage, microbes can contaminate raw coffee in several different ways. The goal of this investigation was to find and isolate contaminants in some coffee powders that were distributed throughout Ghana. For fungus growth, various commercially sold coffee samples gathered from the regions of Ghana were analyzed. With 10% occurrence frequencies, two fungi, *Aspergillus niger* and *Aspergillus flavus*, infected the samples. Additionally, because the raw bean includes the reaction substrates and is processed at extremely high temperatures during roasting, coffee serves as one of the main dietary sources of acrylamide. The benchmark threshold for acrylamide for roasted coffee powder set by the European Commission (400ug/kg) was not detected in any of the commercially available coffee powder samples.

Keywords: commercial coffee powder, microbial, safety, acrylamide, *Aspergillus*

6.1 Introduction

Coffee is one of the most popular beverages in the world and is valued for its sensory qualities, caffeine's stimulant effects, and the bioactive substances it contains, which have positive effects on human health (Schouten et al., 2020). Millions of people around the world depend on it as a key economic crop for their livelihood (Zhang et al., 2019). In 2020/2021, a record 176.1 million kg of coffee was expected to be produced globally, an increase of 5.5 million kg over the previous year (ICO, 2022b). Due to the significant demand and rising sales of coffee beverages globally, the majority of coffee-producing nations strive to

produce a product that is of high quality and safety by implementing food control systems along the production and trade chains.

The two most widely cultivated coffee species are coffee Robusta and coffee Arabica, which together account for 60% and 40% of global production, respectively (Leitão, 2019). Wet and dry processing are the two primary processing techniques utilized at coffee-producing sites to produce intermediate goods. Robusta coffee is often processed in Ghana by dry processing method. Dry Robusta coffee, which is marketed commercially, is roasted and ground into a powder. Coffee is roasted at high temperatures between 200 ^{o}C and 300^{o}C (Cirilo et al., 2003). The most suitable situations for roasting in a laboratory were determined to be 24 minutes at 220-230 ^{o}C for good sensory qualities. The generation of the bioactive and antioxidant chemicals, as well as the main chemical, physical, and organoleptic features of the finished coffee product, are all attributed to the roasting process, which is the most significant unit operation.

In Maillard reactions, reducing sugars (such as glucose and fructose) and asparagine at high temperatures are the main reactants that cause the formation of undesirable acrylamide (Schouten et al., 2020). When foodstuffs acrylamide levels were monitored in Europe during 2007- 2009, it was found that an average of 221ug/kg were present in roasted coffee with a maximum of 2200ug/kg of product (Authority, 2012). When coffee is roasted, the largest amount of acrylamide form at the initial stage and swiftly attain maximum levels. Nevertheless, the longer the roasting continues, the acrylamides levels decrease (Mojska & Gielecinska, 2013). It also appears that the amounts of acrylamide levels may depend on the species,

being that of Arabica or Robusta. When the acrylamide levels of the two species were compared, the latter showed raised levels, in most cases however, these are not statistically significant (Bagdonaite et al., 2008; Lantz et al., 2006).

According to Esquivel and Jimenez (2012), coffee is a tropical product that is widely produced in about 60 tropical and subtropical countries. Coffee farmers encounter numerous issues in the tropics when producing their crops (Harvey et al., 2014). Because coffee is a climate-sensitive plant, the effects of climate change on crop yield and quality as well as the occurrence of invasive pests and fungal diseases have changed how coffee is produced (Pham et al., 2019). A number of pests, illnesses, and fungi affect coffee crops all over the world (Rutherford & Phiri, 2006). In addition to the high humidity, temperature, and rainfall in tropical regions, coffee crops are probably more prone to the development of fungi (Zinedine et al., 2006). Due to the high relative humidity during the time of heavy precipitation, climatic changes can hasten spore formation and fungal production. In Ghana, Robusta coffee is typically fermented using the dry or natural method, which involves removing mature coffee cherries from bushes by hand or machine and spreading them out on the ground for 10 to 25 days. After fermentation, a dry cherry that is devoid of mucilage and pulp but still enclosed by dry skins is left. This cherry is then mechanically stripped of its skins, and beans are then stored there at 11–12% final moisture (Silva et al., 2000). According to Shuping and Eloff (2017), fungal diseases are to blame for the greatest coffee losses worldwide because their diversity is greater than that of bacteria and viruses that cause plant pathology. Two categories of fungi affect coffee: those that affect the crop before

harvest and those that affect the crop after harvest (Brown & Ogle, 1997; Ruffo Roberto et al., 2019).

Pathogenic fungi can infest coffee during all stages of production, including growth, post-harvest handling, storage, and processing (Varga et al., 2010). According to Ismaiel and Papenbrock (2015), nutrient availability, temperature, humidity, and biological factors control post-harvest infections while interactions between plant hosts and other organisms (such as insects) primarily cause fungi infections during the preharvest stage. The vast majority of fungi produce toxins as secondary metabolites, such as ochratoxins-A, the most prevalent mycotoxin found in agricultural products (Lu et al., 2022). This toxin is primarily produced as a result of the secondary metabolism of many species of *Aspergillus* and *Penicillium*. The risk of fungal growth and the production of mycotoxins after harvest is higher in high temperature areas, but toxins-producing fungi can be isolated from coffee beans both before and after harvest (Alvindia & de Guzman, 2016; Daou et al., 2021).

In addition to causing host infections and lowering bean quality, the toxin has been linked to human cancer (Proctor et al., 2018). *Aspergillus*, *Penicillium*, and *Fusarium* are the three main toxigenic fungal genera, and they are all known to naturally contaminate coffee (Bokhari, 2007). Furthermore, according to Rezende et al. (2013), they can infest hosts in both farms and warehouses. Mycotoxins can result in mild to severe side effects like leukopenia, immunodeficiency, and even liver cancer (Perdoncini et al., 2019). According to Fuchs and Peraica (2005) and Reddy and Bhoola (2010), ochratoxin A is classified as a human carcinogen and

teratogen and contains toxic compounds that are harmful to the kidney, liver, and immune system.

Coffee production and quality are both negatively impacted by fungus (Lemessa et al., 2015). Coffee mycotoxins are regarded as a significant food safety concern. One of the most significant mycotoxin pollutants found in agricultural products like cereal, wine, and coffee is ochratoxin-A. The harm it causes to humans and animals, as well as its prevalence, continue to be of concern on a global scale (Ekwomadu et al., 2021). Depending on the level of fungal growth, mycotoxins can be produced. However, the amount of OTA content is significantly reduced during the coffee-roasting, grinding, and brewing processes (Al Attiya et al., 2021; Oliveira et al., 2013; Vieira et al., 2015). Other methods exist for reducing fungi infestation on agricultural and food crops. The use of more effective pre-harvest management techniques, such as crop resistance varieties, crop rotation, appropriate sowing, timings for harvest, and the use of fungicides to control fungal infections, are among them.

Enhancing the quality and safety of coffee beans is essential to prevent the negative issue because some species of fungi can produce mycotoxins, such as ochratoxins and aflatoxins, that are toxic to consumers. Although the fungi can be removed from raw coffee beans by roasting them at an effective temperature, these mycotoxins cannot be removed and would persist in other coffee products, posing a threat to public health. The study sought to evaluate potential fungi pathogens contaminating coffee powder sold in Ghana.

6.2 Materials and Methods

6.2.1 Materials

A total of 100 (10 samples from each region) samples of coffee powder were purchased at random from different retailers throughout the regions of Ghana. They were bought in the major markets of the regions in Ghana. While being transported to the laboratory, the samples were kept in sterile polyethene bags. All samples were kept in a cold room following delivery until use. Samples were labelled according to the regions they were purchased from.

Figure 6.25: Map showing the regions of Ghana where samples were collected

6.2.1.1 Chemical

Hexane and acetonitrile were purchased from Merck in Darmstadt, Germany, and Prolabo VWR International in Paris, France, respectively. Sigma

Aldrich (Germany) provided the salts (MgSO4 and NaCl). The Acros Organics (New Jersey, USA) company provided the acrylamide standard. The Ayensu Starch Company in Ghana's central region supplied analytical starch. The OXOID medium was modified with antibiotic chloramphenicol to make it sterile Potato Dextrose Agar (CM0139).

6.3 Determination of Acrylamide Levels

6.3.1 Extractions and clean-up

This study used a slightly modified sample mass of 2 g of coffee rather than the suggested sample mass of 5 g (Kinsella, 2012). Weighing was done before placing the coffee sample mass, along with the corresponding amounts of $MgSO_4$ (4000 mg) and NaCl (1000 mg), into 50 mL centrifuge tubes. Five (5) mL of hexane were added to the coffee and vortexed for one minute (Wilten and Co. B.V., Holland) to help separate the hydrophilic and hydrophobic components. The mixture was vortexed for an additional minute after the addition of the acetonitrile and distilled water and then centrifuged (LHW 24958, Wageningen) at 3000 rpm for five minutes. After that, 1500 mg of MgSO4 and 500 mg of NaCl were added to the resulting aqueous acetonitrile phase (1 mL), and the mixture was vortexed and agitated at 4000 rpm for five minutes. Finally, 2 cc of the supernatant was taken out for HPLC analysis.

6.3.2 HPLC analysis

According to Gökmen et al. (2005), the HPLC analysis was carried out using a Cecil-Adept binary pump HPLC with a Dynamic Absorbance detector. It

was an Agilent eclipse plus C18 column (4.6 mm 150 mm, 3.5 m) with a column oven set to 25°C. The mobile phase was made up of acetonitrile and water (20:80 v/v), and orthophosphoric acid was used to adjust the pH to 3.5. The flow rate of the mobile phase was set to 1 mL/min, and the wavelength used for detection was 225 nm. For the analysis using the autosampler with both the samples and the standards, a volume of 60 uL was injected into the HPLC. By comparing the acrylamide peak times to the standard retention times, the acrylamide concentrations in the coffee samples were determined. The area beneath the peak times was then automatically integrated by the Cecil-Adept PowerStream (CE 4300, UK).

6.3.3 Quality control

By adding varied quantities of acrylamide standard (20, 50, and 100 g) to 2 g of analytical starch, the method's recovery was examined. The average recovery was 97%, demonstrating the method's sufficiency in terms of accuracy (Chen et al., 2012). A limit of detection (LOD) and limit of quantification (LOQ) of 0.03 g/g and 0.1 g/g, respectively, were employed for the analytical method. This method's calibration was linear and the curve had r^2 of 0.998.

6.4 Extraction and Isolation of Fungi

The simple dilution plate method was adopted in this investigation. Three-fold dilutions were made of 1 g of sample in 9 mL sterilized distilled water. One (1) mL of final dilution (10^{-3}) was flooded on 3 replicates of sterile Potato Dextrose

Agar (PDA) medium that was amended with chloramphenicol antibiotic to suppress bacteria growth.

6.4.1 Identification of fungi

The plates were thereafter incubated at room temperature (25^{o}C-28^{o}C) for 7 days. All fungal growths were counted with a colony counter, and identified with the aid of Illustrated Genera of Imperfect Fungi Manual; authored by Barnett & Hunter (1972). Their frequencies of occurrence were consequently determined.

6.5 Data Analysis

The means and standard deviation of all three replications were analysed using Minitab 16 software. The differences among the test parameters were identified by one-way analysis of variance (ANOVA) using Fisher's least significance differences (LSD) test. All statistical tests were carried out at a 5% significance level.

6.6 Results and Discussion

6.6.1 Acrylamide levels in coffee

Acrylamide has been found in a wide range of foods. Acrylamide levels from the commercially sold coffee in the regions of Ghana ranged from 137.62ug/kg to 328.66ug/kg. Brong Ahafo Region recorded the lowest and the highest was recorded in the Western Region. The differences between the means of acrylamide content of the regions were statistically significant. The significant variation in acrylamide levels between the samples in the current study may be related to the roasting conditions, which may include elements like roasting

temperature, duration, and degree. According to various studies, the amount of acrylamide produced during the early stages of roasting dramatically rises to a maximum point before rapidly falling. As a result, medium-roasted coffee has more acrylamide than dark-roasted coffee. (Mojska & Gielecinska, 2013).

It was possible to visually distinguish differences in the degree of roasting between some of the samples analysed. The variation in acrylamide levels of the studied coffee powders may have been caused by variations in roasting levels, but this may not be the only explanation for the entire variation. When compared to mature coffee beans, coffee beans with defects, specifically immature beans, contain more than two times as much asparagine. As a result, using unmatured coffee beans to make roasted coffee powder may result in higher levels of acrylamide. The types of coffee also affect how much acrylamide is produced during roasting. Asparagine, a limiting element in the production of acrylamide in coffee, is found in higher concentrations in Robusta coffee than in Arabica coffee (Bagdonaite et al., 2008; Lantz et al., 2006).

Coffee powder acrylamide levels are also affected by storage, with levels decreasing over time (Andrzejewski et al., 2004; Soares et al., 2006). The European Commission established 400ug/kg as the benchmark level for acrylamide in roasted coffee powder (Regulation, 2017). All 100 samples of roasted coffee powder obtained from the ten regions of Ghana had acrylamide levels below the benchmark level of 400 ug/kg recommended by the European Commission in Table 6.15.

Table 6.15: Acrylamide levels in commercially sold coffee in Ghana

Samples	Acrylamide content (ug/kg)
AS	258.95±0.55^{c}
BA	137.62±0.59 ^{j}
CR	204.20± 0.47^{f}
ER	156.49± 0.01^{i}
GR	224.39 ± 0.41^{e}
NR	170.53± 0.73 ^{f}
UER	165.69 ± 0.42 ^{h}
UWR	262.12± 0.09^{b}
VR	236.02 ± 0.09^{d}
WR	328.66 ± 0.27^{a}
EC benchmark level	400

Values represent means ± SD of at least three replicates. Means in the same column with the different superscripts are significantly different ($p>0.05$). AS -Ashanti Region, BA – Brong Ahafo Region, CR – Central Region, ER – Eastern Region, GR – Greater Accra Region, NR – Northern Region, UER – Upper East Region, UWR – Upper West Region, VR – Volta Region, WR – Western Region, EC-European commission.

Table 6.16: Acrylamide levels in commercial coffee powder sold in other countries

Countries	Minimum acrylamide content (ug/kg)	Maximum acrylamide content (ug/kg)	Authors
Ethiopia	135	1139	(Deribew & Woldegiorgis, 2021)
Latvia	166	503	(Pugajeva et al., 2015)
Poland	17.7	776.1	(Surma et al., 2017)
Europe	79	1188	(Wenzl & Anklam, 2007)

Researchers from many countries around the world have conducted studies on acrylamide levels of roasted coffee. Deribew and Woldegiorgis (2021) conducted studies with 30 samples of Ethiopia arabica coffee and found the minimum acrylamide levels to be 135ug/kg with a maximum of 1139 ug/kg. In Latia, 22

commercial coffee samples disclosed an acrylamide level ranging from 166ug/kg to 503 ug/kg (Pugajeva et al., 2015). Surma et al. (2017) in Poland used 17 samples of roasted coffee and showed an acrylamide level ranging from 17.7 ug/kg to 776.1 ug/kg. Roasted coffee samples of number 291 analyzed in Europe showed an acrylamide concentration between 79 ug/kg and 1188 ug/kg (Guenther et al., 2007). The minimum acrylamide levels for other researchers were lower than the recorded acrylamide levels in Ghana but the maximum values were higher. The measured levels of acrylamide in commercial coffee powder are below the benchmark (400 ug/kg) set by the European Commission for acrylamide in roasted coffee powder. A long period of storage (Andrzejewski et al., 2004; Soares et al., 2006) and Robusta coffee's roasting conditions (Mojska & Gielecinska, 2013) could be responsible for the lower acrylamide levels.

6.6.2 Mycological analysis

Two contaminating fungi, *Aspergillus niger*, and *Aspergillus flavus* were isolated and recognized in some of the coffee samples bought from the market. *Aspergillus flavus* was isolated and recognized in the coffee sample from the Western Region. Both, *Aspergillus niger* and *Aspergillus flavus* were isolated and identified in the coffee sample from Northern Region. However, no growth was observed in other regions.

Table 6.17: Total fungal count in different coffee powders in the regions

Region of coffee	Fungi isolated	Number isolated	Percentage of contaminated coffee
UER	Negative	0	0
CR	Negative	0	0
WR	*A. flavus*	1	20%
GR	Negative	0	0
BR	Negative	0	0
VR	Negative	0	0
NR	*A. niger* and *A. flavus*	4	80%
AR	Negative	0	0
UWR	Negative	0	0
ER	Negative	0	0
Total		5	100%

Table 6.18: Number of fungi isolated from the commercially sold coffee powder

Fungi	Number isolated
A Niger	2
A flavus	3
Total	5

Mycological analysis showed 20% of *A. flavus* in Western Region roasted coffee, 80% of both *A. niger* and *A. flavus* in Northern Region roasted coffee powder, and 0% in the other regions which is shown in Table 6.18. The highest contamination (80%) was found in roasted coffee from the Northern Region. At least five fungi isolate of different species but the same genera were isolated and most of them belong to food-borne fungi. According to Vega et al. (2008), the nature of sun-drying coffee beans, which is frequently done in open spaces, could be the cause of the fungi isolated from the various products of commercially available powders. Even though mold spores were probably killed off during roasting, some spores are resilient and can remain dormant on a product for a very

long time (Tournas & Katsoudas, 2008). When materials are handled improperly after processing, especially when strict good manufacturing practices and hygienic conditions are not followed, it may be blamed for contaminating the air in the processing area. Among the most widespread fungi in the environment are those that were found in this study. Although it is challenging to completely eradicate them, they can tolerate growth in a variety of substrates and environmental conditions.

In the opinion of Silva et al. (2008), *Aspergillus* is a natural coffee contaminant that can be found from the field to the warehouse. Additionally, due to an unhygienic environment, inadequate heat treatment during roasting or after heat treatment, as well as during packaging, storage, or transportation, could be a contributing factor in the contamination of these fungi in commercial coffee powder. *Aspergillus* in particular, which is known to produce toxic mycotoxins for both humans and animals, is one of the environmental molds whose isolation is of greatest concern. The detrimental metabolites ochratoxin A and aflatoxins are produced by *Aspergillus* (Rahim et al., 2011).

6.7 Conclusion

All of the coffee powder samples analyzed in the current study contained acrylamide. The sample's acrylamide content showed significant variation. The benchmark level for acrylamide for roasted coffee powder set by the European Commission (400ug/kg) was not found in any of the commercially available coffee powder samples. Two distinct species of fungi were also isolated from various types of commercial coffee powder during the mycological analysis. The discovery of

Aspergillus niger and *Aspergillus flavus* suggests unhygienic workers and a poorly maintained manufacturing environment. These fungi can produce toxins and their presence may have an impact on a product's safety. In addition, the product's shelf life is shortened and the flavor of the coffee is diminished. To prevent fungal contamination and ensure the quality of the product, strict good manufacturing practices (GMP) and hygienic procedures should be followed.

CHAPTER SEVEN

PROXIMATE AND MINERAL PARAMETERS OF COMMERCIALLY SOLD COFFEE POWDER

*Vida Gyimah Boadu[1,2], *Ernest Teye[1] & L. K. Sam-Amoah[1]

[1]University of Cape Coast, College of Agriculture and Natural Sciences, School of Agriculture, Department of Agricultural Engineering, Cape Coast, Ghana

[2]Akenten Appiah Menka University of Skills Training and Entrepreneurial Development, Department of Hospitality and Tourism Education, Kumasi Ghana,

*Corresponding authors

Under review: International Journal of Food Science

Statement of contribution of Joint Authorship

Boadu, V. G.: (Candidate)

Conceptualised the topic, established methodology, data collection and analysis, preparation of tables and figures, writing and compilation of the original manuscript.

Teye, E.: (Principal Supervisor)

Conceptualised the topic, established methodology, supervised and edited the manuscript and co-author of manuscript.

Sam-Amoah, L. K.: (Co-Supervisor)

Conceptualised the topic, supervised and edited the manuscript and co-author of manuscript.

Abstract

Coffee is the most commercialized and widely consumed beverage in the world. The reputation of coffee has increased over time due to its preferred flavour and advantageous health effects. Proximate composition and mineral composition analysis of commercially sold coffee powder in the regions of Ghana were investigated. The parameters investigated were determined using standard biochemical methods. The proximate composition of the commercially sold coffee was in the following ranges; moisture (4.48-6.58%), ash (10.45-15.00%), fat (5.85-8.15), protein (9.42-12.94%), and carbohydrate (41.42-52.59%). The mineral composition of coffee was also in the following ranges N (1.50- 2.22%), P (0.35-0.43%), K (1.46-2.05%), Ca (0.46-0.63%), and Mg (0.18-0.26%). The proximate analysis revealed that commercialized coffee powder has a high moisture content which can reduce the shelf life of the product. High ash content was attributed to a substantial quantity of impurities in the coffee samples. The presence of these minerals in coffee beverages depicts a certain nutritional and dietary value when highly consumed.

Keywords: Proximate composition, minerals composition, commercially sold coffee, Robusta

7.1 Introduction

Coffee is one of the popular beverages consumed by millions of people every day. The two species of coffee - coffee arabica (Arabica coffee) and coffee canephora (Robusta) are used to make the majority of coffee beverages consumed worldwide. The coffee beverage, an infusion made from roasted and ground beans,

is primarily responsible for the economic significance of coffee. Its designation as a functional beverage is justified by the impact that coffee consumption provides high content of bioactive compounds with antioxidant in preventing some serious and common ailments.

There are numerous chemical substances found in this beverage, some of which have numerous advantageous properties, and they also function as a stimulant, a function primarily attributable to caffeine. Mood enhancement, improved athletic performance, and a decrease in Parkinson's disease and tremor symptoms are some additional advantages of caffeine (Heckman et al., 2010). Many people make drinking coffee a regular habit and part of their lifestyle. Approximately forty percent of people worldwide commence their days with a cup of coffee (Gaascht et al., 2015). Due mostly to its superior flavour and aroma, coffee is consumed for the health advantages associated with its rich phytochemistry. It has been estimated that coffee contains over a thousand different chemical constituents, including lipids, minerals, nitrogen, alkaloids, and phenolic chemicals (Gaascht et al., 2015; Wachamo, 2017; Zain et al., 2017). Additionally, complex B vitamins, niacin, and chlorogenic acid are present in coffee beans (Belitz et al., 2008).

Like other plant tissues, green coffee beans mostly consist of soluble carbohydrates like fructose, glucose, and galactose as well as insoluble polysaccharides like cellulose and hemicelluloses. To exert their effects, soluble carbohydrates bind fragrance, maintain froth, cause sedimentation, and make the extract more viscous (Arya & Rao, 2007). In addition to these key components,

coffee also contains oils, proteins, and minerals. Coffee beans undergo several chemical changes during the roasting process as a result of the degradation or modification of specific components (Alves et al., 2010), which leads to the emergence of a distinctive aroma, flavour, and colour (Buffo & Cardelli-Freire, 2004). Low molecular weight carbohydrates are produced during roasting by the degradation of polysaccharides. Because of the pyrolysis of organic molecules, the beans' exterior colour (which ranges from light to dark brown) reflects the degree of roast, which affects the aforementioned properties.

Many of the coffee brew's beneficial biological actions are caused by compounds created during roasting (Belitz et al., 2008; Franca & Oliveira, 2009). According to Belitz et al. (2008), roasted coffee contains lipids, proteins, minerals, proteins, and melanoidins. In addition, roasting enhances the quantity of soluble dietary fiber in coffee beans (Silván et al., 2010), and also causes oils to accumulate volatile chemicals that give roasted coffee its distinctive flavour and aroma (Petracco, 2005).

Analysing chemical constituents, physical features, and sensory qualities can help determine the quality of coffee beans. The quality of the beverage may decline as it is processed and stored after leaving the field (Pereira et al., 2021). It is well-recognized that the quality of coffee is correlated with sugars, lipids, and phenolic chemicals (such as caffeine, trigonelline, and chlorogenic acid). The presence of these chemicals has an impact on the price and quality of coffee in the export market. The Maillard reaction and browning, which are responsible for the creation of the colour, flavour, and peculiar aroma of the beverage, are two

important chemical reactions that take place during roasting and are more likely to occur in higher-quality drinking coffee (De Lima et al., 2012). Sucrose, glucose, fructose, arabinose, galactose, and mannose are some of the sugars that are present. Caffeine is one of the components in coffee's chemical makeup that is not destroyed by extensive roasting. In comparison to organic components, which are subject to matter degradation during the roasting process, minerals are also stable (Bitter et al., 2020). During the roasting process for coffee, other compounds like proteins, carbohydrates, chlorogenic acid, and fat may be maintained or destroyed and changed into reactive products (Ginz et al., 2000; Rawel & Kulling, 2007).

Additionally, coffee beverages include dietary sources of micronutrients. According to Janda et al. (2020), the origin, degree of roasting, and brewing technique all affect the concentration of minerals and physiologically active compounds. Among the minerals found in coffee are phosphorus, potassium, magnesium, calcium, sodium, iron, magnesium, zinc, and copper.

The nutrition of humans depends significantly on dietary minerals. Essential minerals are typically divided into macro and microelements depending on the relative amounts in the human body and daily amounts required (greater or less than 100mg/day) (Nabrzyski, 2006). If maintained at acceptable levels, these minerals are essential for cellular integrity, proper nerve function, blood circulation, and bone growth and formation (McDowell, 2003). They also play a role in the manufacture of vitamins, enzymes, and hormones. According to Nieder et al. (2018) and Prashanth et al. (2015), certain nutrients can have detrimental health impacts in excess or shortage. Minerals known as macronutrients are those whose

body weight content surpasses 0.01% and whose demand is higher. There are several of them, including potassium, sodium, calcium, magnesium, phosphorus, sulphur, and chlorine. Less than 0.01% of the body weight of human beings is made up of micronutrients, and their demand is also lower. They include iodine, fluorine, iron, zinc, selenium, copper, manganese, cobalt, and others. In the diet, coffee may provide nutrients, especially if consumed frequently (Janda et al., 2020). Because plants absorb nutrients from the earth, the type of coffee, type of soil, cultivation technique, environmental pollution, and production procedures all have a big impact on the makeup of coffee beans. The most significant factor affecting the mineral makeup of the beans appears to be the provenance of the coffee. The mineral makeup of the soil varies between nations and geographical areas. It is affected by both human influence and the proportions of the naturally existing components (Olechno et al., 2021).

Antioxidants, polyphenols, proximate and mineral composition of green, roasted, defective, and other coffee brands have been reported by other researchers around the world (Adetunji et al., 2021; Degefa et al., 2022; Franca & Oliveira, 2008). Studies was conducted on the proximate composition of raw and roasted Robusta coffee grown in some parts of Ghana (Boadu et al., 2022; Gyedu-Akoto et al., 2019b). The proximate composition and mineral analysis of commercially available coffee in the ten regions of Ghana have not been studied. Investigating the proximate and mineral composition of commercially sold coffee in the regions of Ghana was the primary objective of the current study.

7.2 Materials and Methods

Roasted coffee powders were obtained from major markets in the ten (10) regions of Ghana. A total of 100 coffee samples (10 samples from each region) were purchased randomly from different locations in the market. The samples were kept in sterile polyethene bags during transportation to the laboratory. All the chemicals used in the analysis were of analytical grade. The analysis was conducted at the Crop and soil science Laboratory of the Department of Science and Technology, Kwame Nkrumah University of Science and Technology, Ghana

7.2.1 Proximate determination of commercially sold coffee

A proximate examination was performed on commercially sold coffee. According to AOAC (1990), measurements of moisture, protein, ash, crude fat, and carbohydrate were made.

Protein content. To determine protein content, 1g of each sample was weighed into a Kjeldahl flask along with 25 mL of concentrated sulfuric acid. A catalyst and the sample were digested on a heater for 30 min to produce a clear solution. The solution was cooled using Kjeldahl's method, and then 50mL of a 40 percent NaOH solution was added to neutralize the acid. Following distillation, this was added to a 20 mL drop-mixed indicator and 4% boric acid solution in a 25 mL Erlenmeyer flask. Titrating the solution with 0.1N $H2SO_4$ gave the solution a light pink hue. The nitrogen content was calculated and multiplied by 6.25 to get the protein content.

Fat content. For fat determination, 2.0 g of the dried sample was placed in an extraction thimble and placed inside a Soxhlet apparatus. After being thoroughly cleaned and dried, Soxhlet flasks were filled with weighed petroleum ether (60^{o}C). After assembly, the extractors were refluxed for 6 h. After the ether had been allowed to evaporate in a rotary vapourizer, the flasks were removed, cooled, and weighed. The weight of the extracted fat was determined as a portion of the sample weight.

Moisture content. Moisture content was determined by weighing 5.0g of each sample into porcelain crucibles and dried in an oven at a constant weight at 105^{o}C. The percentage of the samples' total weight was used to calculate the amount of moisture lost during drying.

Ash content. About 2g of each sample was weighed and put into porcelain crucibles and burned in a muffle furnace for 4 hours at 550^{o}C to generate carbon-free ash. This was used to determine how much ash was present. The weight of the ash was calculated as a proportion of the sample weight.

Carbohydrate content. Differences from 100% given by equation (7.1) in moisture, protein, fat, and ash yielded carbohydrates. Equation (7.2) was used to compute the energy.

% Carbohydrate = 100 – (moisture + protein +ash + fat) ……. Equation (7.1)

Energy (Kcal/100g) = (Protein x 9) + (Fat x 4) + (Carbohydrate x 4) ……. Equation (7.2)

7.2.2 Mineral determination of commercially sold coffee

Atomic Absorption Spectroscopy (ASS) was used to determine the minerals (N, P, K, Ca, and Mg) content after samples were digested per AOAC, 1990 with modifications. Each flour sample was weighed into Kjeldahl digestion tubes at a rate of 0.25 g, and then 7.5 mL of concentrated H_2SO_4 and 2.5 mL of concentrated HNO_3 were added. The materials were broken down at 300°C for 4 h until the colour of the solution was cleared. The clear solution was diluted with 50 mL of distilled water after being cooled to room temperature. Before the mineral components could be measured with an atomic absorption spectrophotometer, the diluted solution was warmed to evaporate and atomize them.

7.2.3 Data analysis

The means and standard deviation of all three replications were analysed using Minitab 16 software. The differences among the test parameters were identified by one-way analysis of variance (ANOVA) using Fisher's least significance differences (LSD) test. All statistical tests were carried out at a 5% significance level.

7.3 Results and discussion

7.3.1 Proximate composition of commercially sold coffee

The moisture content of the commercially sold powder was high ranging from 4.84% and 6.58% for the samples from the ten regions as shown in Table 7.19. The moisture content of all the samples was higher than values recorded by (Vasconcelos et al., 2007) for light (0.9%), medium (0.9%), and dark roast (1.0%)

coffee values and 3.7% recorded in literature (Gyedu-Akoto et al., 2019b). It was lower than the 14.6% recorded by the Danish Food Composition Databank (DFCD, 2009). The moisture content is used to determine how susceptible and microbiological stable as well as to quantify the water activity of the food (Adetunji et al., 2021). High-moisture food products have a shorter shelf life because they are more vulnerable to microbial attack and deterioration (Hassan & Umar, 2004).

Table 7.19: Proximate Analysis of commercially sold coffee

Samples	Moisture (%)	Ash (%)	Fat (%)	Protein (%)	CHO (%)	CME (kCal/ 100g)
AS	5.96±0.63a	13.80±2.26ab	7.50±2.12a	9.42±1.39b	52.59±1.15a	303.54
BA	6.15±0.14a	14.30±2.55ab	7.20±0.99a	11.95±0.48ab	43.59±1.53ab	286.96
CR	5.75±0.18a	14.85±0.49ab	7.80±0.14a	11.51±1.17ab	41.71±0.54ab	295.08
ER	5.91±0.89a	15.00±0.84a	8.15±2.33a	12.76±0.59a	41.42±2.23b	290.07
GR	5.43±0.36a	14.65±1.63ab	6.70±1.41a	11.59±0.20ab	45.74±3.05ab	289.62
NR	4.84±0.40a	10.55±2.62b	6.10±1.27a	13.79±0.67a	49.84±2.17a	309.42
UER	5.12±0.10a	12.20±2.83ab	5.85±0.49a	12.87±0.62a	47.63±4.63ab	294.65
UWR	6.52±0.19a	12.30±1.56ab	7.65±1.20a	11.17±0.30ab	48.30±2.07a	306.73
VR	5.47±0.50a	14.15±1.77ab	6.90±2.26a	12.92±0.55a	48.91±4.42a	309.42
WR	6.58±1.96a	10.45±0.070b	7.50±1.84a	12.94±0.651a	46.31±3.51ab	304.50
DFCD	14.6	4.0	5.5	15.4	60	322

Values represent means ± SD of at least three replicates. Means in the same column with the different superscripts are significantly different ($p>0.05$). AS -Ashanti Region, BA – Brong Ahafo Region, CR – Central Region, ER – Eastern Region, GR – Greater Accra Region, NR – Northern Region, UER – Upper East Region, UWR – Upper West Region, VR – Volta Region, WR – Western Region, CHO – Carbohydrate, CME - Calculated Metabolic Energy, DFCD – Danish Food Composition Databank

According to Marshall (2010), the ash content of food represents its overall mineral composition. The ash concentration of the coffee powder ranged from 10.45% to 15.00%, with the commercially coffee sold in Eastern region having the highest ash content and the Western region having the lowest. Except for the Western and Northern regions, the recorded ash levels were greater than the

literature values for (Vasconcelos et al., 2007) in the range of (4.4- 4.6%) and for (Gyedu-Akoto et al., 2019b) in the value of 4.4%. It was also higher than the 4.0% recorded by (DFCD, 2009). Coffee is cultivated in six out of 10 regions in Ghana namely Ashanti, Brong Ahafo, Eastern, Central, Western and Volta region (Wongnaa et al., 2021). Commercially powder coffee is processed and transported to neighboring regions where coffee is not cultivated. The differences in coffee plantations and their location could account for the changes in ash content in the regions where coffee is grown (Oliveira et al., 2013). Higher ash content could also be attributed to a substantial quantity of impurities in the roasted coffee powder samples (Muller et al., 2013). Some of the impurities may be husks, sticks, leaves defective coffee, foreign materials, and other unknown adulterants.

The fat concentration ranged from 5.20% in the Upper East region to 8.15% in the Eastern region. The crude fat recorded for the regions was all lower than the fat value (12.2%) recorded by (Gyedu-Akoto et al., 2019b) but in the range (5.5%) with (DFCD, 2009). Fat in coffee serves as a carrier for flavors and fat-soluble vitamins and contributes to the texture and mouthfeel of the brew (Oestreich, 2010). Low fat in samples may decrease the flavour, and change the texture and mouthfeel of the coffee. The crude protein content of the roasted coffee samples ranged from 9.42% to 13.79% in (Table 7.19) with the Ashanti region recording the lowest (9.42%) whilst the Northern region recorded the highest protein (13.79%). Protein values recorded were all below the ranges stated by (Vasconcelos et al., 2007) (13.1-13.2%), (DFCD, 2009) (15.4) and higher than (Gyedu-Akoto et al., 2019b) value (5.6%).

The carbohydrate content of samples was in a range from 41.42% to 52.59%. All the recorded values are lower than the values in the literature (71.1-72.9%) for (Vasconcelos et al., 2007), 60% for (DFCD, 2009) and 74.2% (Gyedu-Akoto et al., 2019b). Due to the various roasting techniques, the amount of carbohydrates and proteins was low. It has been found that the temperature and time of roasting have an impact on the chemical composition, aroma and colour of coffee beans (Buffo & Cardelli-Freire, 2004). Proteins, peptides, and amino acids are crucial for the quality of coffee cups. They interact with carbohydrates during roasting to produce flavors and aromas known as the Maillard reaction (Mazzafera et al., 2019).

7.3.2 Mineral composition of commercially sold coffee

The mineral analysis results of commercially sold coffee are shown in Table 8.2. Nitrogen (N) content ranged from 1.50% to 2.22%. The highest level of Nitrogen was found in the sample for the Northern region and the lowest was found in the Ashanti region. In the case of Phosphorus (P), the highest content thereof was found in the Western region and the lowest was found in the Ashanti region. It was in the range of 0.35% to 0.43%. Commercially sold coffee from Greater Accra was the richest in potassium (K) with a value of (2.03%) and the lowest was recorded in the Upper West region of value (1.46%). There was no significant difference among the values recorded for phosphorus and potassium. Calcium (Ca) was in the range of 0.46% to 0.63% while Magnesium (Mg) was also ranging from 0.18% to 0.26%. This supports claims that many elements, including geography (water, soil, flora and climate), agronomic products, post-harvest management, and other stress

factors, affect the mineral content of coffee beans (Gyedu-Akoto et al., 2019b). In light of the extremely high consumption of coffee worldwide, it has been stated that the presence of major and minor mineral elements in coffee and its derivatives is crucial since they demonstrate a particular nutritional and dietary value (Pohl et al., 2013). As a result, coffee mineral analysis is important for determining daily intakes of various minerals and trace elements.

Table 7.18: Mineral composition (%) of commercially sold coffee

Samples	N/%	P/%	K/%	Ca/%	Mg/%
AS	1.50±0.53^{b}	0.37±0.01^{a}	1.58±0.07^{a}	0.63±0.01^{a}	0.26±0.00^{a}
BA	1.91±0.07^{ab}	0.39±0.07^{a}	1.83±0.07^{a}	0.53±0.06^{bc}	0.20±0.044^{bc}
CR	1.86±0.15^{ab}	0.35±0.04^{a}	1.98±0.00^{a}	0.51±0.03^{bc}	0.22±0.05^{abc}
ER	2.05±0.08^{a}	0.42±0.00^{a}	1.83±0.24^{a}	0.51±0.01^{bc}	0.22±0.02^{abc}
GR	1.85±0.03^{ab}	0.38±0.04^{a}	2.03±0.07^{a}	0.63±0.02^{a}	0.22±0.01^{abc}
NR	2.22±0.08^{a}	0.39±0.02^{a}	1.51±0.67^{a}	0.49±0.01^{bc}	0.19±0.02^{bc}
UER	2.06±0.09^{a}	0.38±0.07^{a}	2.05±0.20^{a}	0.55±0.03^{ab}	0.20±0.01^{bc}
UWR	1.81±0.07^{ab}	0.36±0.06^{a}	1.46±0.60^{a}	0.51±0.02^{bc}	0.21±0.00^{abc}
VR	2.07±0.08^{a}	0.40±0.02^{a}	1.99±0.18^{a}	0.54±0.03^{b}	0.25±0.01^{ab}
WR	2.08±0.10^{a}	0.43±0.00^{a}	1.59±0.08^{a}	0.46±0.05^{c}	0.18±0.03^{c}

Values represent means ± SD of at least three replicates. Means in the same column with the different superscripts are significantly different ($p>0.05$). AS -Ashanti Region, BA – Brong Ahafo Region, CR – Central Region, ER – Eastern Region, GR – Greater Accra Region, NR – Northern Region, UER – Upper East Region, UWR – Upper West Region, VR – Volta Region, WR – Western Region

7.4 Conclusion

The proximate analysis revealed that commercialized coffee powder had high moisture and ash content. The high moisture content can reduce the shelf life and quality parameters of the coffee powder. High ash content was attributed to a significant quantity of impurities (coffee husk, sticks, leaves, foreign materials, and other coffee adulterants) in the roasted coffee powder samples. A significant amount of these minerals (N, P, K, Ca, Mg) were present in the coffee samples. The

presence of minerals in coffee beverages depicts a certain nutritional and dietary value when consumed frequently.

CHAPTER EIGHT

SUMMARY, GENERAL CONCLUSIONS AND RECOMMENDATION

8.1 Summary

The research study sought to authenticate different forms of coffee varieties and African geographical coffee types (raw, roasted, and powder) by using pocket-sized NIR spectroscopy and multivariate data modelling. Two varieties (Arabica and Robusta) and their forms (raw, roasted and powder) were discriminated by a handheld spectrometer. Also, African Robusta coffee and its coffee types were identified by the handheld spectrometer. Another aim of this work was to detect adulteration of coffee husk in roasted coffee powder by NIR spectroscopic method. Coffee beans were bought, roasted, and processed into powder. This was adulterated with coffee husk powder at different concentrations to assess the onsite application of a handheld NIR spectrometer to detect adulteration. Furthermore, commercially sold coffee powder was purchased from the major local markets in Ghana for quality and fungus determination. Microbial, acrylamide, proximate and mineral analysis were conducted on the samples.

8.1.1 Key findings

The following findings regarding the research objectives that guided the study were revealed concerning the first research objective. NIR spectroscopy in the range of 740-1070nm provided some important information in the study. For the authentication of coffee bean varieties (Arabica and Robusta) of different forms (raw, roasted, and powder), the best results obtained for the following were as follows; raw coffee beans SD-SVM had an accuracy of 0.92 and efficiency of 0.82.

For roasted coffee beans, SD-KNN had an accuracy of 0.92 and an efficiency of 0.87 while for roasted powdered coffee; FD-KNN showed an accuracy of 0.97 and an efficiency of 0.97. These findings reveal that for more accurate differentiation of coffee beans, the roasted powder offers the best results.

The second objective was to discriminate African coffee types using NIR spectroscopy in the above range and the results were as follows: The best classification algorithms were developed for the following coffee types: raw bean coffee, SD-PLSDA, and MC+SD-PLSDA. These models had an accuracy of 0.87 and an F1-score of 0.88. SNV+SD-SVM and MSC+SD-NN both had accuracy and F1 scores of 0.97 for roasted coffee beans and 0.96 for roasted coffee powder, respectively. These results demonstrate that roasted coffee beans generate the best results for a more precise classification of coffee beans.

The third objective sought to detect coffee husks in Robusta coffee. Authentic roasted powder coffee was adulterated with coffee husk in the following percentages; 30%, 25%, 20%, 15%, 10%, 5%, and 0% representing the pure unadulterated sample. The results showed that in the qualitative detection of adulterants, FD-LDA model had 97.78 % and 100 % in both the calibration set and prediction set. While SPA-PLS model had R = 0.9711 and 0.9897 in both the calibration set and prediction set in the quantitative detection of adulterants. The outcome of this study showed short-wave NIR spectroscopy could be used for coffee integrity examination.

The fourth objective sought to determine the microbial quality and acrylamide levels of commercially sold coffee from the major local markets in

Ghana. With 10% occurrence frequencies, two fungi, *Aspergillus niger* and *Aspergillus flavus*, infected the samples. The Northern Region of Ghana recorded the highest with 80% contaminated coffee with Western Region recording 20% contaminated coffee. There was no contamination in the rest of the major markets in Ghana. Additionally, because the raw bean includes the reaction substrates and is processed at extremely high temperatures during roasting. Coffee serves as one of the main dietary sources of acrylamide. The benchmark threshold for acrylamide for roasted coffee powder set by the European Commission (400ug/kg) was not detected in any of the commercially available coffee powder samples. All the values recorded were below the benchmark threshold.

The fifth objective was to determine the proximate and mineral composition of the commercially sold coffee. The proximate composition of the commercially sold coffee was in the following ranges: moisture (4.48-6.58%), ash (10.45-15.00%), fat (5.85-8.15), protein (9.42-12.94%), and carbohydrate (41.42-52.59%). The mineral composition of coffee was also in the following ranges N (1.50-2.22%), P (0.35-0.43%), K (1.46-2.05%), Ca (0.46-0.63%), and Mg (0.18-0.26%). The proximate analysis revealed that commercialized coffee powder has a high moisture content which will reduce the shelf life of the product. High ash content was attributed to a substantial quantity of impurities in the coffee samples. The presence of these minerals in coffee beverages depicts a certain nutritional and dietary value when highly consumed.

8.2 Conclusion

Findings from the first objective suggest that pocket-size near-infrared spectroscopy coupled with appropriate chemometric analysis is a rapid and non-destructive method for authenticating different categories of coffee variety. The methodical selection of different pre-processing techniques (FD, SD, MC, MSC, and SNV) with PCA modelling together with PLSDA, KNN and SVM presented an advantage in authenticating the coffee varieties in three states (green, roasted and powder). Generally, it could be concluded that pocket size spectroscopy together with chemometrics could be used to exploit rapid authenticating of different varieties of coffee. Besides, these models can be imported into smartphones for effective and accurate authenticating of coffee.

In the second objective, geographical differentiation of African Robusta coffee types (bean, roasted, powdered) could be accomplished using handheld near-infrared spectroscopy in combination with an appropriate multivariant classification algorithm in chemometrics. Accuracy and the F1-score were used to evaluate the model's effectiveness. According to the results, it is possible to geographically differentiate between various types of African coffee using analytical data from handheld NIR spectroscopy. This technique could benefit coffee bean producers in developing countries like; Ghana, Uganda, Burkina Faso, Ivory Coast and processors mostly in industrialized countries like Japan, Holland, and Germany.

In addition to requiring trained personnel, destruction, and the use of pricey

reagents, using chemicals to solely detect adulteration is difficult and time-consuming. However, a quick, on-site, and inexpensive method for coffee examination could be used, combining short-wave NIR spectroscopy with chemometrics. First derivative (FD) pre-processing modelled with linear discriminate analysis (FD-LDA), which focuses on quality, was found to be the best technique. On the other hand, the SPA-PLS model performed the best in both the calibration set and the prediction set for the quantitative detection of coffee adulterants in genuine coffee.

In the determination of microbial and quality parameters of commercially sold coffee, acrylamide was present in all the analyzed samples. The sample's acrylamide content showed significant variation. The acrylamide levels were all below the benchmark level for roasted coffee powder set by the European Commission (400ug/kg). Two distinct species of fungi were also isolated from various types of commercial coffee powder during the mycological analysis. The discovery of Aspergillus Niger and Aspergillus flavus suggests unhygienic workers and a poorly maintained manufacturing environment. These fungi can produce toxins and their presence may have an impact on a product's safety. In addition, the product's shelf life is shortened and the flavor of the coffee is diminished. To prevent fungal contamination and ensure the quality of the product, strict GMP and hygienic procedures should be followed.

In the fifth objective, the proximate analysis revealed that commercialized coffee powder had high moisture (4.84 -6.58%) and ash content (10.45% - 15.00%). The high moisture content will reduce the shelf life and quality parameters of the

coffee powder. High ash content was attributed to a significant quantity of impurities (coffee husk, sticks, leaves, foreign materials, and other coffee adulterants) in the roasted coffee powder samples. A significant amount of these minerals (N, P, K, Ca, Mg) were present in the coffee samples. The presence of minerals in coffee beverages depicts a certain nutritional and dietary value when consumed frequently.

8.3 Recommendation

In this study, the potential and adaptability of portable near-infrared spectroscopy combined with chemometrics methods were demonstrated for the analysis of coffee beans. This study has demonstrated that this environmentally friendly, quick analytical technique can handle specialized analyses in the field of coffee bean spectroscopy for both qualitative and quantitative determinations. On the other hand, developing a stable and reliable calibration model for handheld NIR spectroscopy is very challenging and calls for a large number of prior measurements for NIR method development, and as a result, it must be done very carefully. Additionally, it is challenging to derive accurate, reliable models because they require many samples to account for a wide range of variations.

Chemometric methods have made it extremely simple to develop a stable and reliable model for analyzing coffee beans, but it's nonetheless essential to choose the right approach for deriving information from NIR spectroscopy data. As a result, numerous studies ought to concentrate on a variety of chemometrics technique combinations.

Further study should be focused on the feasibility of handheld NIR spectroscopy for the authenticating regional and geographical origins of coffee beans since it was done in only five countries. Again, additional research is required to confirm the performance across different locations and varieties in detecting adulterants. Handheld NIR spectroscopy alone could sometimes not give optimum performance. Therefore, the upcoming trends should include the fusion of handheld NIR spectrometers with imaging techniques to improve the analysis of trivial components of quantity and quality and concerns such as the prediction of chemical residue, adulterants, and mycotoxins in coffee beans.

Furthermore, even though handheld NIR spectroscopy can replace wet analytical chemistry methods, the main coffee-growing nations lack sufficient income and technical know-how to operate them effectively. These challenges make it difficult to use them in everyday situations. Therefore, the development of affordable handheld spectrometers must go hand in hand with intensive training in handheld NIR spectroscopy for the major players in the coffee industry. Furthermore, more research should be done over the coming years to confirm and establish the usefulness of handheld NIR spectroscopy for the analysis of coffee beans.

Lastly, research on acrylamide levels, fungi development, proximate composition, and shelf life of commercially available coffee powder in other areas and countries should be carried out.

References

Abbas, O., Pissard, A., & Baeten, V. (2020). Near-infrared, mid-infrared, and Raman spectroscopy. In *Chemical Analysis of Food* (pp. 77-134). Elsevier.

Adetunji, O. A., Coker-Osiwoga, T. F., Okwuanalu, C., Olusola, M., Adetayo, O. C. A., Ogunbiyi, B. T., Amah, G. H., & Adetunji, O. A. (2021). Comparative Study of Phytochemical Constituents, Caffeine Levels and Proximate Composition of Liven Alkaline Coffee, Nescafe Original Coffee and Nescafe Original Decaffeinated Coffee. EAS J Nutr Food Sci, 3, 94-99.

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11).

Adnan, A., von Hörsten, D., Pawelzik, E., & Mörlein, D. (2017). Rapid prediction of moisture content in intact green coffee beans using near infrared spectroscopy. *Foods*, *6*(5), 38.

Agresti, P. D. M., Franca, A. S., Oliveira, L. S., & Augusti, R. (2008). Discrimination between defective and non-defective Brazilian coffee beans by their volatile profile. *Food Chemistry*, *106*(2), 787-796.

Aguilera, Y., Rebollo-Hernanz, M., Cañas, S., Taladrid, D., & Martín-Cabrejas, M. A. (2019). Response surface methodology to optimise the heat-assisted aqueous extraction of phenolic compounds from coffee parchment and their comprehensive analysis. *Food & Function*, *10*(8), 4739-4750.

Al Attiya, W., Hassan, Z. U., Al-Thani, R., & Jaoua, S. (2021). Prevalence of toxigenic fungi and mycotoxins in Arabic coffee (Coffea arabica):

Protective role of traditional coffee roasting, brewing and bacterial volatiles. *PloS one*, *16*(10), e0259302.

Alessandrini, L., Romani, S., Pinnavaia, G., & Dalla Rosa, M. (2008). Near infrared spectroscopy: An analytical tool to predict coffee roasting degree. *Analytica Chimica Acta*, *625*(1), 95-102.

Alves, R. C., Almeida, I. M., Casal, S., & Oliveira, M. B. P. (2010). Isoflavones in coffee: influence of species, roast degree, and brewing method. *Journal of Agricultural and Food Chemistry*, *58*(5), 3002-3007.

Alves, R. C., Rodrigues, F., Nunes, M. A., Vinha, A. F., & Oliveira, M. B. P. (2017). State of the art in coffee processing by-products. *Handbook of coffee processing by-products*, 1-26.

Alvindia, D. G., & de Guzman, M. F. (2016). Survey of Philippine coffee beans for the presence of ochratoxigenic fungi. *Mycotoxin Research*, *32*(2), 61-67.

Amalia, F., Aditiawati, P., Putri, S. P., & Fukusaki, E. (2021). Gas chromatography/mass spectrometry-based metabolite profiling of coffee beans obtained from different altitudes and origins with various postharvest processing. *Metabolomics*, *17*(7), 1-16.

Amorim Filho, V. R., Polito, W. L., & Gomes Neto, J. A. (2007). Comparative studies of the sample decomposition of green and roasted coffee for determination of nutrients and data exploratory analysis. *Journal of the Brazilian Chemical Society*, *18*, 47-53.

Anderson, K. A., & Smith, B. W. (2002). Chemical profiling to differentiate geographic growing origins of coffee. *Journal of Agricultural and Food Chemistry*, *50*(7), 2068-2075.

Andrzejewski, D., Roach, J. A., Gay, M. L., & Musser, S. M. (2004). Analysis of coffee for the presence of acrylamide by LC-MS/MS. *Journal of Agricultural and Food Chemistry*, *52*(7), 1996-2002.

Anyidoho, E. K., Teye, E., & Agbemafle, R. (2020). Nondestructive authentication of the regional and geographical origin of cocoa beans by using a handheld NIR spectrometer and multivariate algorithm. *Analytical Methods*, *12*(33), 4150-4158.

AOAC. (1990). Official methods of analysis. In: Aoac Washington, DC.

AOAC. (1995). Official methods of analysis, 16 edn. AOAC. *Washington*.

Arrieta, A. A., Arrieta, P. L., & Mendoza, J. M. (2019). Analysis of coffee adulterated with roasted corn and roasted soybean using voltammetric electronic tongue. *Acta Scientiarum Polonorum Technologia Alimentaria*, *18*(1), 35-41.

Arya, M., & Rao, L. J. M. (2007). An impression of coffee carbohydrates. *Critical Reviews in Food Science and Nutrition*, *47*(1), 51-67.

Authority, E. F. S. (2012). Update on acrylamide levels in food from monitoring years 2007 to 2010. *EFSA Journal*, *10*(10), 2938.

Ayu, P., Budiastra, I., & Rindang, A. (2020). NIR spectroscopy application for determination caffeine content of Arabica green bean coffee. IOP Conference Series: Earth and Environmental Science, 454(1), 012049. IOP Publishing.

Aziz, F., Rahman, M. M., Ahmad, T., Jahan, M. S., Tosrif, T. M., Huq, M. M., Reza, C., Ahmed, A. U., Day, P., & Badsha, S. (2014). Discrimination analysis of EEG signals at eye open and eye close condition for ECS

switching system. 2013 International Conference on Electrical Information and Communication Technology (EICT), 1-4, IEEE.

Bagdonaite, K., Derler, K., & Murkovic, M. (2008). Determination of acrylamide during roasting of coffee. *Journal of Agricultural and Food Chemistry*, *56*(15), 6081-6086.

Baggenstoss, J., Perren, R., & Escher, F. (2008). Water content of roasted coffee: impact on grinding behaviour, extraction, and aroma retention. *European Food Research and Technology*, *227*, 1357-1365.

Baker, M. (2016). Arabica vs. Robusta comparison. Turkish Style Coffee. In.

Bansal, S., Singh, A., Mangal, M., Mangal, A. K., & Kumar, S. (2017). Food adulteration: Sources, health risks, and detection methods. *Critical Reviews in Food Science and Nutrition*, *57*(6), 1174-1189.

Baptistao, M., de Carvalho Rocha, W. F., & Poppi, R. J. (2011). Quality control of the paracetamol drug by chemometrics and imaging spectroscopy in the near infrared region. *Journal of Molecular Structure*, *1002*(1-3), 167-171.

Baqueta, M. R., Coqueiro, A., Março, P. H., Mandrone, M., Poli, F., & Valderrama, P. (2021). Integrated 1H NMR fingerprint with NIR spectroscopy, sensory properties, and quality parameters in a multi-block data analysis using ComDim to evaluate coffee blends. *Food Chemistry*, *355*, 129618.

Barbin, D. F., Felicio, A. L. d. S. M., Sun, D.-W., Nixdorf, S. L., & Hirooka, E. Y. (2014). Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview. *Food Research International*, *61*, 23-32.

Barboza, F. D., & Poppi, R. J. (2003). Determination of alcohol content in beverages using short-wave near-infrared spectroscopy and temperature

correction by transfer calibration procedures. *Analytical and Bioanalytical Chemistry*, *377*, 695-701.

Barnes, R., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, *43*(5), 772-777.

Barnett, H. L., & Hunter, B. B. (1972). Illustrated genera of imperfect fungi. *Illustrated genera of imperfect fungi.*, (3rd ed), 241.

Batista, L. R., Chalfoun, S. M., Prado, G., Schwan, R. F., & Wheals, A. E. (2003). Toxigenic fungi associated with processed (green) coffee beans (*Coffea arabica* L.). *International Journal of Food Microbiology*, *85*(3), 293-300.

Beć, K. B., Grabska, J., & Huck, C. W. (2021). Principles and applications of miniaturized near-infrared (NIR) spectrometers. *Chemistry–A European Journal*, *27*(5), 1514-1532.

Becalski, A., Lau, B. P.-Y., Lewis, D., & Seaman, S. W. (2003). Acrylamide in foods: occurrence, sources, and modeling. *Journal of Agricultural and Food Chemistry*, *51*(3), 802-808.

Bekalo, S. A., & Reinhardt, H.-W. (2010). Fibers of coffee husk and hulls for the production of particleboard. *Materials and Structures*, *43*(8), 1049-1060.

Belitz, H.-D., Grosch, W., & Schieberle, P. (2008). *Food chemistry*. Springer Science & Business Media.

Bermudez, S., Vivek Voora, V., & Larrea, C. (2022). Coffee prices and sustainability. *Sustainable Commodities Marketplace Series*.

Bi, Y., Yuan, K., Xiao, W., Wu, J., Shi, C., Xia, J., Chu, G., Zhang, G., & Zhou, G. (2016). A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. *Analytica Chimica Acta*, *909*, 30-40.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197-227.

Bicchi, C. P., Panero, O. M., Pellegrino, G. M., & Vanni, A. C. (1997). Characterization of roasted coffee and coffee beverages by solid phase microextraction− gas chromatography and principal component analysis. *Journal of Agricultural and Food Chemistry*, *45*(12), 4680-4686.

Bilge, G. (2020). Investigating the effects of geographical origin, roasting degree, particle size and brewing method on the physicochemical and spectral properties of Arabica coffee by PCA analysis. *Journal of Food Science and Technology*, *57*(9), 3345-3354.

Bitter, N. Q., Fernandez, D. P., Driscoll, A. W., Howa, J. D., & Ehleringer, J. R. (2020). Distinguishing the region-of-origin of roasted coffee beans with trace element ratios. *Food Chemistry*, *320*, 126602.

Blanco, M., & Villarroya, I. (2002). NIR spectroscopy: a rapid-response analytical tool. *TrAC Trends in Analytical Chemistry*, *21*(4), 240-250.

Boadu, V. G., Teye, E., Amuah, C. L., Lamptey, F. P., & Sam-Amoah, L. K. (2023). Portable NIR Spectroscopic Application for Coffee Integrity and Detection of Adulteration with Coffee Husk. *Processes*, *11*(4), 1140.

Boadu, V. G., Teye, E., Amuah, C. L., & Sam-Amoah, L. (2022). Rapid authentication of coffee bean varieties of different forms by using a pocket-

sized spectrometer and multivariate data modelling. *Analytical Methods*, *14*(46), 4756-4766.

Bondesson, E. (2015). A nutritional analysis on the by-product coffee husk and its potential utilization in food production.

Botelho, B. G., Mendes, B. A., & Sena, M. M. (2013). Implementação de um método robusto para o controle fiscal de umidade em queijo minas artesanal. Abordagem metrológica multivariada. *Química Nova*, *36*, 1416-1422.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Bressanello, D., Liberto, E., Cordero, C., Rubiolo, P., Pellegrino, G., Ruosi, M. R., & Bicchi, C. (2017). Coffee aroma: Chemometric comparison of the chemical information provided by three different samplings combined with GC–MS to describe the sensory properties in cup. *Food Chemistry*, *214*, 218-226.

Briandet, R., Kemsley, E. K., & Wilson, R. H. (1996). Discrimination of Arabica and Robusta in instant coffee by Fourier transform infrared spectroscopy and chemometrics. *Journal of Agricultural and Food Chemistry*, *44*(1), 170-174.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical methods*, *6*(9), 2812-2831.

Brown, J. F., & Ogle, H. (1997). *Plant pathogens and plant diseases*. Published by Rockvale Publications for the Division of Botany, University of New England.

Budiastra, I., Widyotomo, S., & Ayu, P.C (2018). Prediction of caffeine content in java preanger coffee beans by NIR spectroscopy using PLS and MLR method. IOP conference series: earth and environmental science, 147(1), 012004, IOP Publishing.

Budryn, G., & Nebesny, E. (2008). Antioxidant properties of Arabica and Robusta coffee extracts prepared under different conditions. *Deutsche Lebensmittel-Rundschau, 104*(2), 69-78.

Buffo, R. A., & Cardelli-Freire, C. (2004). Coffee flavour: an overview. *Flavour and Fragrance Journal*, *19*(2), 99-104.

Buratti, S., Sinelli, N., Bertone, E., Venturello, A., Casiraghi, E., & Geobaldo, F. (2015). Discrimination between washed Arabica, natural Arabica and Robusta coffees by using near infrared spectroscopy, electronic nose and electronic tongue analysis. *Journal of the Science of Food and Agriculture*, *95*(11), 2192-2200.

Cagliani, L. R., Pellegrino, G., Giugno, G., & Consonni, R. (2013). Quantification of Coffea arabica and Coffea canephora var. robusta in roasted and ground coffee blends. *Talanta*, *106*, 169-173.

Cämmerer, B., & Kroh, L. W. (2006). Antioxidant activity of coffee brews. *European Food Research and Technology*, *223*(4), 469-474.

Candolfi, A., De Maesschalck, R., Jouan-Rimbaud, D., Hailey, P., & Massart, D. (1999). The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *Journal of Pharmaceutical and Biomedical Analysis*, *21*(1), 115-132.

Cano-Marquina, A., Tarín, J., & Cano, A. (2013). The impact of coffee on health. *Maturitas*, *75*(1), 7-21.

Cao, D.-S., Huang, J.-H., Liang, Y.-Z., Xu, Q.-S., & Zhang, L.-X. (2012). Tree-based ensemble methods and their applications in analytical chemistry. *TrAC Trends in Analytical Chemistry*, *40*, 158-167.

Casal, S., Oliveira, M., Alves, M., & Ferreira, M. (2000). Discriminate analysis of roasted coffee varieties for trigonelline, nicotinic acid, and caffeine content. *Journal of Agricultural and Food Chemistry*, *48*(8), 3420-3424.

Catelani, T. A., Páscoa, R. N., Santos, J. R., Pezza, L., Pezza, H. R., Lima, J. L., & Lopes, J. A. (2017). A non-invasive real-time methodology for the quantification of antioxidant properties in coffee during the roasting process based on near-infrared spectroscopy. *Food and Bioprocess Technology*, *10*, 630-638.

Catelani, T. A., Santos, J. R., Páscoa, R. N., Pezza, L., Pezza, H. R., & Lopes, J. A. (2018). Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. *Talanta*, *179*, 292-299.

Cen, H., & He, Y. (2007). Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science & Technology*, *18*(2), 72-83.

Chakravartula, S. S. N., Moscetti, R., Bedini, G., Nardella, M., & Massantini, R. (2022). Use of convolutional neural network (CNN) combined with FT-NIR

spectroscopy to predict food adulteration: A case study on coffee. *Food Control*, *135*, 108816.

Chandrasekar, V., & Viswanathan, R. (1999). Physical and thermal properties of coffee. *Journal of Agricultural Engineering Research*, *73*(3), 227-234.

Chen, L., Xue, X., Ye, Z., Zhou, J., Chen, F., & Zhao, J. (2011). Determination of Chinese honey adulterated with high fructose corn syrup by near infrared spectroscopy. *Food Chemistry*, *128*(4), 1110-1114.

Chen, M.-J., Hsu, H.-T., Lin, C.-L., & Ju, W.-Y. (2012). A statistical regression model for the estimation of acrylamide concentrations in French fries for excess lifetime cancer risk assessment. *Food and Chemical Toxicology*, *50*(10), 3867-3876.

Chen, Q., Cai, J., Wan, X., & Zhao, J. (2011). Application of linear/non-linear classification algorithms in discrimination of pork storage time using Fourier transform near infrared (FT-NIR) spectroscopy. *LWT-Food Science and Technology*, *44*(10), 2053-2058.

Chen, Q., Chen, M., Liu, Y., Wu, J., Wang, X., Ouyang, Q., & Chen, X. (2018). Application of FT-NIR spectroscopy for simultaneous estimation of taste quality and taste-related compounds content of black tea. *Journal of food Science and Technology*, *55*, 4363-4368.

Chen, X., Sun, X., Hua, H., Yi, Y., Li, H., & Chen, C. (2019). Quality evaluation of decoction pieces of Rhizoma Atractylodis Macrocephalae by near infrared spectroscopy coupled with chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *221*, 117169.

Choi, E., Choi, K.-H., Park, S. M., Shin, D., Joh, H.-K., & Cho, E. (2016). The benefit of bone health by drinking coffee among Korean postmenopausal women: A cross-sectional analysis of the fourth & fifth Korea national health and nutrition examination surveys. *PloS one*, *11*(1), e0147762.

Christin, C., Hoefsloot, H. C., Smilde, A. K., Hoekman, B., Suits, F., Bischoff, R., & Horvatovich, P. (2013). A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. *Molecular & Cellular Proteomics*, *12*(1), 263-276.

Chu, Y.-F. (2012). *Coffee: emerging health effects and disease prevention* (Vol. 59). John Wiley & Sons.

Cirilo, M. P., Coelho, A. F. S., Araújo, C. M., Gonçalves, F. R., Nogueira, F. D., & Glória, M. B. A. (2003). Profile and levels of bioactive amines in green and roasted coffee. *Food Chemistry*, *82*(3), 397-402.

Ciza, P., Sacre, P.-Y., Waffo, C., Coïc, L., Avohou, H., Mbinze, J., Ngono, R., Marini, R., Hubert, P., & Ziemons, E. (2019). Comparing the qualitative performances of handheld NIR and Raman spectrophotometers for the detection of falsified pharmaceutical products. *Talanta*, *202*, 469-478.

Clark, H. A., & Snedeker, S. M. (2006). Ochratoxin A: its cancer risk and potential for exposure. *Journal of Toxicology and Environmental Health, Part B: Critical Reviews*, *9*(3), 265-296.

Clarke, R. J. (2012). *Coffee: Volume 2: Technology* (Vol. 2). Springer Science & Business Media.

Conway, J. (2020). Coffee Consumption Worldwide from 2012/13 to 2020/21.

Coronel-Reyes, J., Ramirez-Morales, I., Fernandez-Blanco, E., Rivero, D., & Pazos, A. (2018). Determination of egg storage time at room temperature using a low-cost NIR spectrometer and machine learning techniques. *Computers and Electronics in Agriculture*, *145*, 1-10.

Correia, R. M., Andrade, R., Tosato, F., Nascimento, M. T., Pereira, L. L., Araújo, J. B., Pinto, F. E., Endringer, D. C., Padovan, M. P., & Castro, E. V. (2020). Analysis of Robusta coffee cultivated in agroforestry systems (AFS) by ESI-FT-ICR MS and portable NIR associated with sensory analysis. *Journal of Food Composition and Analysis*, *94*, 103637.

Correia, R. M., Tosato, F., Domingos, E., Rodrigues, R. R., Aquino, L. F. M., Filgueiras, P. R., Lacerda Jr, V., & Romao, W. (2018). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta*, *176*, 59-68.

Correia, R. M., Tosato, F., Domingos, E., Rodrigues, R. R., Aquino, L. F. M., Filgueiras, P. R., Lacerda Jr, V., & Romao, W. (2018). Portable near infrared spectroscopy applied to quality control of Brazilian coffee. *Talanta*, *176*, 59-68.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297.

Craig, A. P., Franca, A. S., & Oliveira, L. S. (2012). Discrimination between defective and non-defective roasted coffees by diffuse reflectance infrared Fourier transform spectroscopy. *Lwt*, *47*(2), 505-511.

Daglia, M., Papetti, A., Gregotti, C., Bertè, F., & Gazzani, G. (2000). In vitro antioxidant and ex vivo protective activities of green and roasted coffee. *Journal of Agricultural and Food Chemistry*, *48*(5), 1449-1454.

Dalianis, H. (2018). Evaluation metrics and evaluation. In *Clinical text mining* (pp. 45-53). Springer.

DaMatta, F. M., Ronchi, C. P., Maestri, M., & Barros, R. S. (2007). Ecophysiology of coffee growth and production. *Brazilian Journal of Plant Physiology*, *19*, 485-510.

Danish Food Composition Databank (DFCD) (2009), Ed. 7.01, FCDB no. 0103, National Food Institute – Technical University of Denmark (DTU).

Dankowska, A., Domagała, A., & Kowalewski, W. (2017). Quantification of Coffea arabica and Coffea canephora var. robusta concentration in blends by means of synchronous fluorescence and UV-Vis spectroscopies. *Talanta*, *172*, 215-220.

Daou, R., Joubrane, K., Maroun, R. G., Khabbaz, L. R., Ismail, A., & El Khoury, A. (2021). Mycotoxins: Factors influencing production and control strategies. *AIMS Agriculture and Food*, *6*(1), 416-447.

Davis, A., Govaerts, R., Bridson, D., & Stoffelen, P. (2006). An annotated taxonomic conspectus of the genus. In: Coffea.

de Carvalho Couto, C., Freitas-Silva, O., Morais Oliveira, E. M., Sousa, C., & Casal, S. (2021). Near-Infrared Spectroscopy Applied to the Detection of Multiple Adulterants in Roasted and Ground Arabica Coffee. *Foods*, *11*(1), 61.

De Lima, L. M., Pozza, E. A., & Da Silva Santos, F. (2012). Relationship between incidence of brown eye spot of coffee cherries and the chemical composition of coffee beans. *Journal of Phytopathology*, *160*(4), 209-211.

De Matteis, R., Arch, J., Petroni, M., Ferrari, D., Cinti, S., & Stock, M. (2002). Immunohistochemical identification of the β3-adrenoceptor in intact human adipocytes and ventricular myocardium: effect of obesity and treatment with ephedrine and caffeine. *International Journal of Obesity*, *26*(11), 1442-1450.

de Melo Pereira, G. V., de Carvalho Neto, D. P., Júnior, A. I. M., do Prado, F. G., Pagnoncelli, M. G. B., Karp, S. G., & Soccol, C. R. (2020). Chemical composition and health properties of coffee and coffee by-products. *Advances in Food and Nutrition Research*, *91*, 65-96.

Dégardin, K., Guillemain, A., & Roggo, Y. (2017). Comprehensive study of a handheld Raman spectrometer for the analysis of counterfeits of solid-dosage form medicines. *Journal of Spectroscopy*, *2017*.

Degefa, M., Alamerew, S., Mohammed, A., & Gemechu, A. (2022). Biochemical Composition Variation among Southern Ethiopian Arabica Coffee (Coffea arabica L.) Genotypes. *International Journal of Agronomy*, *2022*.

Deribew, H. A., & Woldegiorgis, A. Z. (2021). Acrylamide levels in coffee powder, potato chips and French fries in Addis Ababa city of Ethiopia. *Food Control*, *123*, 107727.

Dhanoa, M., Lister, S., Sanderson, R., & Barnes, R. (1994). The link between multiplicative scatter correction (MSC) and standard normal variate (SNV)

transformations of NIR spectra. *Journal of Near Infrared Spectroscopy*, *2*(1), 43-47.

Dharmawan, A., Masithoh, R. E., & Amanah, H. Z. (2023). Development of PCA-MLP Model Based on Visible and Shortwave Near Infrared Spectroscopy for Authenticating Arabica Coffee Origins. *Foods*, *12*(11), 2112.

Doran, M., Raicu, D. S., Furst, J. D., Settimi, R., Schipma, M., & Chandler, D. P. (2007). Oligonucleotide microarray identification of Bacillus anthracis strains using support vector machines. *Bioinformatics*, *23*(4), 487-492.

Dórea, J. G., & da Costa, T. H. M. (2005). Is coffee a functional food? *British Journal of Nutrition*, *93*(6), 773-782.

Downey, G., & Boussion, J. (1996). Authentication of coffee bean variety by near-infrared reflectance spectroscopy of dried extract. *Journal of the Science of Food and Agriculture*, *71*(1), 41-49.

Duckworth, J. (2004). Mathematical data preprocessing. *Near-infrared Spectroscopy in Agriculture*, *44*, 113-132.

Dybkowska, E., Sadowska, A., Rakowska, R., Debowska, M., Swiderski, F., & Swiader, K. (2017). Assessing polyphenols content and antioxidant activity in coffee beans according to origin and the degree of roasting. *Roczniki Państwowego Zakładu Higieny*, *68*(4), 347-353.

Ebrahimi-Najafabadi, H., Leardi, R., Oliveri, P., Casolino, M. C., Jalali-Heravi, M., & Lanteri, S. (2012). Detection of addition of barley to coffee using near infrared spectroscopy and chemometric techniques. *Talanta*, *99*, 175-179.

Eira, M. T., Silva, E., De Castro, R. D., Dussert, S., Walters, C., Bewley, J. D., & Hilhorst, H. W. (2006). Coffee seed physiology. *Brazilian Journal of Plant Physiology, 18*, 149-163.

Ekwomadu, T. I., Akinola, S. A., & Mwanza, M. (2021). Fusarium mycotoxins, their metabolites (Free, emerging, and masked), food safety concerns, and health impacts. *International Journal of Environmental Research and Public Health*, *18*(22), 11741.

Endaye, M., Atlabachew, M., Mehari, B., Alemayehu, M., Mengistu, D. A., & Kerisew, B. (2020). Combining multi-element analysis with statistical modeling for tracing the origin of green coffee beans from Amhara region, Ethiopia. *Biological Trace Element Research*, *195*(2), 669-678.

Esquivel, P., & Jimenez, V. M. (2012). Functional properties of coffee and coffee by-products. *Food Research International*, *46*(2), 488-495.

Esteban-Dıez, I., González-Sáiz, J., & Pizarro, C. (2004). An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS. *Analytica Chimica Acta*, *514*(1), 57-67.

Esteban-Díez, I., González-Sáiz, J., Sáenz-González, C., & Pizarro, C. (2007). Coffee varietal differentiation based on near infrared spectroscopy. *Talanta*, *71*(1), 221-229.

Esteban-Díez, I., González-Sáiz, J., Sáenz-González, C., & Pizarro, C. (2007). Coffee varietal differentiation based on near infrared spectroscopy. *Talanta*, *71*(1), 221-229.

Esteki, M., Shahsavari, Z., & Simal-Gandara, J. (2018). Use of spectroscopic methods in combination with linear discriminant analysis for authentication of food products. *Food Control*, *91*, 100-112.

FAO. (2022). Market and trade: coffee. *Food and Agriculture Organisation of United Nations*.

Farah, A., & dos Santos, T. F. (2015). The coffee plant and beans: An introduction. In *Coffee in health and disease prevention* (pp. 5-10). Elsevier.

Farah, A., Monteiro, M. C., Calado, V., Franca, A. S., & Trugo, L. (2006). Correlation between cup quality and chemical attributes of Brazilian coffee. *Food Chemistry*, *98*(2), 373-380.

Fraga, C. G., Galleano, M., Verstraeten, S. V., & Oteiza, P. I. (2010). Basic biochemical mechanisms behind the health benefits of polyphenols. *Molecular Aspects of Medicine*, *31*(6), 435-445.

Franca, A. S., & Oliveira, L. S. (2008). Chemistry of defective coffee beans. *Food chemistry research Developments*, *4*, 105-138.

Franca, A. S., & Oliveira, L. S. (2009). Coffee processing solid wastes: current uses and future perspectives. *Agricultural Wastes*, *9*, 155-189.

Franca, A. S., Oliveira, L. S., Mendonça, J. C., & Silva, X. A. (2005). Physical and chemical attributes of defective crude and roasted coffee beans. *Food chemistry*, *90*(1-2), 89-94.

Fuchs, R., & Peraica, M. (2005). Ochratoxin A in human kidney diseases. *Food additives and contaminants*, *22*(s1), 53-57.

Gaascht, F., Dicato, M., & Diederich, M. (2015). Coffee provides a natural multitarget pharmacopeia against the hallmarks of cancer. *Genes & nutrition*, *10*, 1-17.

Gałuszka, A., Migaszewski, Z. M., & Namieśnik, J. (2015). Moving your laboratories to the field–Advantages and limitations of the use of field portable instruments in environmental sample analysis. *Environmental Research*, *140*, 593-603.

Garattini, S. (1993). *Caffeine, coffee, and health*. Raven Press.

Gil-Agusti, M. T., Campostrini, N., Zolla, L., Ciambella, C., Invernizzi, C., & Righetti, P. G. (2005). Two-dimensional mapping as a tool for classification of green coffee bean species. *Proteomics*, *5*(3), 710-718.

Ginz, M., Balzer, H. H., Bradbury, A. G., & Maier, H. G. (2000). Formation of aliphatic acids by carbohydrate degradation during roasting of coffee. *European Food Research and Technology*, *211*, 404-410.

Giraudo, A., Grassi, S., Savorani, F., Gavoci, G., Casiraghi, E., & Geobaldo, F. (2019). Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis. *Food Control*, *99*, 137-145.

Gökmen, V., Şenyuva, H. Z., Acar, J., & Sarıoğlu, K. (2005). Determination of acrylamide in potato chips and crisps by high-performance liquid chromatography. *Journal of Chromatography A*, *1088*(1-2), 193-199.

González, A., Pablos, F., Martín, M., León-Camacho, M., & Valdenebro, M. (2001). HPLC analysis of tocopherols and triglycerides in coffee and their use as authentication parameters. *Food Chemistry*, *73*(1), 93-101.

Gonzalez-Rios, O., Suarez-Quiroz, M. L., Boulanger, R., Barel, M., Guyot, B., Guiraud, J.-P., & Schorr-Galindo, S. (2007). Impact of "ecological" post-harvest processing on coffee aroma: II. Roasted coffee. *Journal of Food Composition and Analysis*, *20*(3-4), 297-307.

Gorji, H. T., Shahabi, S. M., Sharma, A., Tande, L. Q., Husarik, K., Qin, J., Chan, D. E., Baek, I., Kim, M. S., & MacKinnon, N. (2022). Combining deep learning and fluorescence imaging to automatically identify fecal contamination on meat carcasses. *Scientific Reports*, *12*(1), 2392.

Gouvea, B., Torres, C., Franca, A., Oliveira, L., & Oliveira, E. (2009). Feasibility of ethanol production from coffee husks. *Biotechnology Letters*, *31*(9), 1315-1319.

Guenther, H., Anklam, E., Wenzl, T., & Stadler, R. H. (2007). Acrylamide in coffee: review of progress in analysis, formation and level reduction. *Food Additives and Contaminants*, *24*(sup1), 60-70.

Gunasekaran, S., & Irudayaraj, J. (2000). Optical methods—Visible, NIR and FTIR spectroscopy. *Nondestructive Food Evaluation—Techniques to Analyze Properties and Quality*, 1-38.

Gyedu-Akoto, E., Opoku, S. Y., & Ofosu-Agyei, E. (2019). Proximate, mineral composition and sensory evaluation of coffee and kola flavoured biscuits. *Nutrition & Food Science*, *49*(4), 752-761.

Gyedu-Akoto, E., Opoku, S. Y., & Ofosu-Agyei, E. (2019). Proximate, mineral composition and sensory evaluation of coffee and kola flavoured biscuits. *Nutrition & Food Science*, *49*(4), 752-761.

Harvey, C. A., Rakotobe, Z. L., Rao, N. S., Dave, R., Razafimahatratra, H., Rabarijohn, R. H., Rajaofara, H., & MacKinnon, J. L. (2014). Extreme vulnerability of smallholder farmers to agricultural risks and climate change in Madagascar. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1639), 20130089.

Hassan, L., & Umar, K. (2004). Proximate and mineral composition of seeds and pulp of African locust bean (Parkia biglobosa L.). *Nigerian Journal of Basic and Applied Sciences*, *13*, 15-27.

Hečimović, I., Belščak-Cvitanović, A., Horžić, D., & Komes, D. (2011). Comparative study of polyphenols and caffeine in different coffee varieties affected by the degree of roasting. *Food Chemistry*, *129*(3), 991-1000.

Heckman, M. A., Weil, J., & De Mejia, E. G. (2010). Caffeine (1, 3, 7-trimethylxanthine) in foods: a comprehensive review on consumption, functionality, safety, and regulatory matters. *Journal of Food Science*, *75*(3), R77-R87.

Huck, C., Guggenbichler, W., & Bonn, G. (2005). Analysis of caffeine, theobromine and theophylline in coffee by near infrared spectroscopy (NIRS) compared to high-performance liquid chromatography (HPLC) coupled to mass spectrometry. *Analytica Chimica Acta*, *538*(1-2), 195-203.

ICO. (2009). International Coffee Organization). *ICO trade statistics, exports by exporting countries to all destinations.www.ico.org/trade_statistics.asp*.

ICO. (2012). International Coffee Organization. In: Monthly Coffee Market Report.

ICO. (2018). Coffee production by exporting countries. *International Coffee Organisation*.

ICO. (2021). International Coffee Organization.

ICO. (2022). Coffee Market Report.

ICO. (2022). Daily Coffee Prices. In I. C. Organisation (Ed.), *Internation Coffee Organisation*.

Inácio, M. R. C., de Lima, K. M. G., Lopes, V. G., Pessoa, J. D. C., & de Almeida Teixeira, G. H. (2013). Total anthocyanin content determination in intact açaí (Euterpe oleracea Mart.) and palmitero-juçara (Euterpe edulis Mart.) fruit using near infrared spectroscopy (NIR) and multivariate calibration. *Food Chemistry*, *136*(3-4), 1160-1164.

Ismaiel, A. A., & Papenbrock, J. (2015). Mycotoxins: producing fungi and mechanisms of phytotoxicity. *Agriculture*, *5*(3), 492-537.

Janda, K., Jakubczyk, K., Baranowska-Bosiacka, I., Kapczuk, P., Kochman, J., Rębacz-Maron, E., & Gutowska, I. (2020). Mineral composition and antioxidant potential of coffee beverages depending on the brewing method. *Foods*, *9*(2), 121.

Janissen, B., & Huynh, T. (2018). Chemical composition and value-adding applications of coffee industry by-products: A review. *Resources, Conservation and Recycling*, *128*, 110-117.

Jiao, Y., Hall, J. J., & Morton, Y. T. (2017). Automatic equatorial GPS amplitude scintillation detection using a machine learning algorithm. *IEEE Transactions on Aerospace and Electronic Systems*, *53*(1), 405-418.

Jing, M., Cai, W., & Shao, X. (2010). Quantitative determination of the components in corn and tobacco samples by using near-infrared spectroscopy and multiblock partial least squares. *Analytical Letters*, *43*(12), 1910-1921.

Joint, F., & Additives, W. E. C. o. F. (2001). *Safety evaluation of certain mycotoxins in food*. FAO.

Kademi, H. I., Ulusoy, B. H., & Hecer, C. (2019). Applications of miniaturized and portable near infrared spectroscopy (NIRS) for inspection and control of meat and meat products. *Food Reviews International*, *35*(3), 201-220.

Kamiloglu, S. (2019). Authenticity and traceability in beverages. *Food chemistry*, *277*, 12-24.

Kamruzzaman, M., Sun, D.-W., ElMasry, G., & Allen, P. (2013). Fast detection and visualization of minced lamb meat adulteration using NIR hyperspectral imaging and multivariate image analysis. *Talanta*, *103*, 130-136.

Kemsley, E. K., Ruault, S., & Wilson, R. H. (1995). Discrimination between Coffea arabica and Coffea canephora variant robusta beans using infrared spectroscopy. *Food Chemistry*, *54*(3), 321-326.

Khamar, K. (2013). Short text classification using kNN based on distance function. *International Journal of advanced research in computer and communication engineering*, *2*(4), 1916-1919.

Kinsella, B. (2012). Acrylamide by QuEChERS extraction with LC-MS-MS detection. *LC-GC North America*, *30*(9), S35-S35.

Kljusurić, J. G., Mihalev, K., Bečić, I., Polović, I., Georgieva, M., Djaković, S., & Kurtanjek, Ž. (2016). Near-infrared spectroscopic analysis of total phenolic content and antioxidant activity of berry fruits. *Food Technology and Biotechnology*, *54*(2), 236.

Kranenburg, R. F., Ramaker, H. J., Sap, S., & van Asten, A. C. (2022). A calibration friendly approach to identify drugs of abuse mixtures with a portable near-infrared analyzer. *Drug Testing and Analysis*, *14*(6), 1089-1101.

Krishnan, S. (2017). Sustainable coffee production. In *Oxford research encyclopedia of environmental science*.

Kufa, T. (2010). Environmental sustainability and coffee diversity in Africa. ICO world coffee conference,

Lantz, I., Ternité, R., Wilkens, J., Hoenicke, K., Guenther, H., & van der Stegen, G. H. (2006). Studies on acrylamide levels in roasting, storage and brewing of coffee. *Molecular Nutrition & Food Research*, *50*(11), 1039-1046.

Leardi, R., & Nørgaard, L. (2004). Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. *Journal of Chemometrics: A Journal of the Chemometrics Society*, *18*(11), 486-497.

Lee, L. C., & Jemain, A. A. (2019). Predictive modelling of colossal ATR-FTIR spectral data using PLS-DA: empirical differences between PLS1-DA and PLS2-DA algorithms. *Analyst*, *144*(8), 2670-2678.

Lee, L. C., Liong, C.-Y., & Jemain, A. A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst*, *143*(15), 3526-3539.

Lee, L. C., Liong, C.-Y., & Jemain, A. A. (2018). Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst*, *143*(15), 3526-3539.

Leitão, A. L. (2019). Occurrence of ochratoxin A in coffee: Threads and solutions—A mini-review. *Beverages*, *5*(2), 36.

Lemessa, F., Abera, A., Adunga, G., & Garedew, W. (2015). Association of mycoflora with coffee (Coffea arabica L.) beans at Limmu coffee plantation, Southwestern Ethiopia. *Plant Pathology Journal*, *14*(3), 136.

Levate Macedo, L., da Silva Araújo, C., Costa Vimercati, W., Gherardi Hein, P. R., Pimenta, C. J., & Henriques Saraiva, S. (2021). Evaluation of chemical properties of intact green coffee beans using near-infrared spectroscopy. *Journal of the Science of Food and Agriculture*, *101*(8), 3500-3507.

Li, J., Zhao, C., Huang, W., Zhang, C., & Peng, Y. (2014). A combination algorithm for variable selection to determine soluble solid content and firmness of pears. *Analytical Methods*, *6*(7), 2170-2180.

Liu, Z., Zuo, M. J., Zhao, X., & Xu, H. (2015). An Analytical Approach to Fast Parameter Selection of Gaussian RBF Kernel for Support Vector Machine. *J. Inf. Sci. Eng.*, *31*(2), 691-710.

Lu, L., Tibpromma, S., Karunarathna, S. C., Jayawardena, R. S., Lumyong, S., Xu, J., & Hyde, K. D. (2022). Comprehensive review of fungi on coffee. *Pathogens*, *11*(4), 411.

Luna, A. S., da Silva, A. P., Pinho, J. S., Ferré, J., & Boqué, R. (2013). Rapid characterization of transgenic and non-transgenic soybean oils by chemometric methods using NIR spectroscopy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *100*, 115-119.

Lyman, D. J., Benck, R., Dell, S., Merle, S., & Murray-Wijelath, J. (2003). FTIR-ATR analysis of brewed coffee: effect of roasting conditions. *Journal of Agricultural and Food Chemistry*, *51*(11), 3268-3272.

Manach, C., Scalbert, A., Morand, C., Rémésy, C., & Jiménez, L. (2004). Polyphenols: food sources and bioavailability. *The American Journal of Clinical Nutrition*, *79*(5), 727-747.

Mancini, M., Mazzoni, L., Gagliardi, F., Balducci, F., Duca, D., Toscano, G., Mezzetti, B., & Capocasa, F. (2020). Application of the non-destructive NIR technique for the evaluation of strawberry fruits quality parameters. *Foods*, *9*(4), 441.

Manuel, M. N. B., da Silva, A. C., Lopes, G. S., & Ribeiro, L. P. D. (2022). One-class classification of special agroforestry Brazilian coffee using NIR spectrometry and chemometric tools. *Food Chemistry*, *366*, 130480.

Marquetti, I., Link, J. V., Lemes, A. L. G., dos Santos Scholz, M. B., Valderrama, P., & Bona, E. (2016). Partial least square with discriminant analysis and near infrared spectroscopy for evaluation of geographic and genotypic origin of arabica coffee. *Computers and Electronics in Agriculture, 121*, 313-319.

Marshall, M. R. (2010). Ash analysis. *Food Analysis, 4*, 105-116.

Martín, M. a. J., Pablos, F., González, A. G., Valdenebro, M. a. S., & León-Camacho, M. (2001). Fatty acid profiles as discriminant parameters for coffee varieties differentiation. *Talanta, 54*(2), 291-297.

Martın, M., Pablos, F., & González, A. (1999). Characterization of arabica and robusta roasted coffee varieties and mixture resolution according to their metal content. *Food Chemistry, 66*(3), 365-370.

Martini, D., Del Bo', C., Tassotti, M., Riso, P., Del Rio, D., Brighenti, F., & Porrini, M. (2016). Coffee consumption and oxidative stress: a review of human intervention studies. *Molecules, 21*(8), 979.

Massart, D. L., Vandeginste, B. G., Buydens, L., Lewi, P., Smeyers-Verbeke, J., & Jong, S. D. (1998). *Handbook of Chemometrics and Qualimetrics*. Elsevier Science Inc.

Mattila, P., Hellström, J., & Törrönen, R. (2006). Phenolic acids in berries, fruits, and beverages. *Journal of Agricultural and Food Chemistry, 54*(19), 7193-7199.

Mazzafera, P., Schimpl, F., & Kiyota, E. (2019). Proteins of coffee beans: Recent advances. *Coffee*, 429-444.

McDowell, L. R. (2003). *Minerals in animal and human nutrition*. Academic Press Inc.

Mees, C., Souard, F., Delporte, C., Deconinck, E., Stoffelen, P., Stévigny, C., Kauffmann, J.-M., & De Braekeleer, K. (2018). Identification of coffee leaves using FT-NIR spectroscopy and SIMCA. *Talanta*, *177*, 4-11.

Meilgaard, M. C., Carr, B. T., & Civille, G. V. (1999). *Sensory evaluation techniques*. CRC press.

Mendes, L. C., de Menezes, H. C., Aparecida, M., & Da Silva, A. (2001). Optimization of the roasting of robusta coffee (C. canephora conillon) using acceptability tests and RSM. *Food Quality and Preference*, *12*(2), 153-162.

Mendonça, J. C., Franca, A. S., Oliveira, L. S., & Afonso, R. (2009). Application of electrospray ionization-mass spectrometry to the discrimination of green and roasted coffees by species and quality. *Handbook on mass spectrometry: instrumentation, data and analysis, and applications. New York: Nova Publishers*, *17*.

Mendonça, J. C., Franca, A. S., Oliveira, L. S., & Nunes, M. (2008). Chemical characterisation of non-defective and defective green arabica and robusta coffees by electrospray ionization-mass spectrometry (ESI-MS). *Food Chemistry*, *111*(2), 490-497.

Mesias, M., Delgado-Andrade, C., & Morales, F. J. (2022). An updated view of acrylamide in cereal products. *Current Opinion in Food Science*, *46*, 100847.

Miao, X., Miao, Y., Gong, H., Tao, S., Chen, Z., Wang, J., Chen, Y., & Chen, Y. (2021). NIR spectroscopy coupled with chemometric algorithms for the prediction of cadmium content in rice samples. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *257*, 119700.

Mojska, H., & Gielecinska, I. (2013). Studies of acrylamide level in coffee and coffee substitutes: influence of raw material and manufacturing conditions. *Roczniki Państwowego Zakładu Higieny*, *64*(3).

Mondal, T. K., & Das, S. (2014, December). Discrimination of various Milk adulterated sample using Linear Discrimination Analysis (LDA). In *2014 Annual IEEE India Conference (INDICON)* (pp. 1-4). IEEE.

Moon, J.-K., & Shibamoto, T. (2009). Role of roasting conditions in the profile of volatile flavor chemicals formed from coffee beans. *Journal of Agricultural and Food Chemistry*, *57*(13), 5823-5831.

Morales, F. J., Martin, S., Açar, Ö. Ç., Arribas-Lorenzo, G., & Gökmen, V. (2009). Antioxidant activity of cookies and its relationship with heat-processing contaminants: a risk/benefit approach. *European Food Research and Technology*, *228*, 345-354.

Morgano, M. A., Faria, C. G., Ferrão, M. F., Bragagnolo, N., & Ferreira, M. M. d. C. (2008). Determination of moisture in raw coffee by near infra-red reflectance spectroscopy and multivariate regression. *Food Science and Technology*, *28*, 12-17.

Mottram, D. S., Wedzicha, B. L., & Dodson, A. T. (2002). Acrylamide is formed in the Maillard reaction. *Nature*, *419*(6906), 448-449.

Muller, A., Huebner, L., & Souza, C. (2013). Evaluation of physico-chemical quality of different brands of roasted coffee soluble and in powder marketed in the region of Vale do Taquari/RS. *Revista Brasileira de Tecnologia Agroindustrial*, *7*, 1004-1012.

Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2002). An Introduction to Kernel-Based Learning Algorithms. In *Handbook of neural network signal processing: Neural network signal processing* (pp. 95-134). CRC Press.

Murthy, P. S., & Naidu, M. M. (2012). Sustainable management of coffee industry by-products and value addition—A review. *Resources, Conservation and Recycling*, *66*, 45-58.

Mussatto, S. I., Machado, E. M., Martins, S., & Teixeira, J. A. (2011). Production, composition, and application of coffee and its industrial residues. *Food and Bioprocess Technology*, *4*, 661-672.

Mutz, Y. S., do Rosario, D., Galvan, D., Schwan, R. F., Bernardes, P. C., & Conte-Junior, C. A. (2023). Feasibility of NIR spectroscopy coupled with chemometrics for classification of Brazilian specialty coffee. *Food Control*, 109696.

Nabais, J. M. V., Nunes, P., Carrott, P. J., Carrott, M. M. L. R., García, A. M., & Díaz-Díez, M. (2008). Production of activated carbons from coffee endocarp by CO2 and steam activation. *Fuel Processing Technology*, *89*(3), 262-268.

Nabrzyski, M. (2006). Functional role of some minerals in foods. In *Mineral Components in Foods* (pp. 123-161). CRC Press.

Nescatelli, R., Carradori, S., Marini, F., Caponigro, V., Bucci, R., De Monte, C., Mollica, A., Mannina, L., Ceruso, M., & Supuran, C. T. (2017). Geographical characterization by MAE-HPLC and NIR methodologies and carbonic anhydrase inhibition of Saffron components. *Food Chemistry*, *221*, 855-863.

Nestle, S. (2004). Faces of coffee: the Nestle coffee report.

Nieder, R., Benbi, D. K., Reichl, F. X., Nieder, R., Benbi, D. K., & Reichl, F. X. (2018). Macro-and secondary elements and their role in human health. *Soil Components and Human H ealth*, 257-315.

Nørgaard, L. (2005). iToolbox Manual. In: FOSS, Data & Algorithms in R&D.

Norgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Applied spectroscopy*, *54*(3), 413-419.

Nunes, F. M., & Coimbra, M. A. (2001). Chemical characterization of the high molecular weight material extracted with hot water from green and roasted arabica coffee. *Journal of Agricultural and Food Chemistry*, *49*(4), 1773-1782.

Núñez, N., Martínez, C., Saurina, J., & Núñez, O. (2021). High-performance liquid chromatography with fluorescence detection fingerprints as chemical descriptors to authenticate the origin, variety and roasting degree of coffee

by multivariate chemometric methods. *Journal of the Science of Food and Agriculture*, *101*(1), 65-73.

Núñez, N., Saurina, J., & Núñez, O. (2021). Non-targeted HPLC-FLD fingerprinting for the detection and quantitation of adulterated coffee samples by chemometrics. *Food Control*, *124*, 107912.

Oestreich, J. S. (2010). Chemistry of coffee. *Comprehensive Natural Products. Oxford: Elsevier*, *2*(3), 1085-1117.

Oestreich-Janzen, S. (2013). Chemistry of Coffee. *Elsevier Inc.*, 1-24.

Okaru, A. O., Scharinger, A., Rajcic de Rezende, T., Teipel, J., Kuballa, T., Walch, S. G., & Lachenmeier, D. W. (2020). Validation of a quantitative proton nuclear magnetic resonance spectroscopic screening method for coffee quality and authenticity (NMR Coffee Screener). *Foods*, *9*(1), 47.

Okubo, N., & Kurata, Y. (2019). Nondestructive classification analysis of green coffee beans by using near-infrared spectroscopy. *Foods*, *8*(2), 82.

Olechno, E., Puścion-Jakubik, A., Socha, K., & Zujko, M. E. (2021). Coffee brews: Are they a source of macroelements in human nutrition? *Foods*, *10*(6), 1328.

Oliveira, G., da Silva, D. M., Pereira, R. G. F. A., Paiva, L. C., Prado, G., & Batista, L. R. (2013). Effect of different roasting levels and particle sizes on ochratoxin A concentration in coffee beans. *Food Control*, *34*(2), 651-656.

Oliveira, L. S., Franca, A. S., Mendonça, J. C., & Barros-Júnior, M. C. (2006). Proximate composition and fatty acids profile of green and roasted defective coffee beans. *LWT-Food Science and Technology*, *39*(3), 235-239.

Oliveira, M., Casal, S., Morais, S., Alves, C., Dias, F., Ramos, S., Mendes, E., Delerue-Matos, C., & Oliveira, M. B. P. (2012). Intra-and interspecific mineral composition variability of commercial instant coffees and coffee substitutes: Contribution to mineral intake. *Food Chemistry*, *130*(3), 702-709.

Oliveira, R. C., Oliveira, L. S., Franca, A. S., & Augusti, R. (2009). Evaluation of the potential of SPME-GC-MS and chemometrics to detect adulteration of ground roasted coffee with roasted barley. *Journal of Food Composition and Analysis*, *22*(3), 257-261.

Oliveira, R. C., Oliveira, L. S., Franca, A. S., & Augusti, R. (2009). Evaluation of the potential of SPME-GC-MS and chemometrics to detect adulteration of ground roasted coffee with roasted barley. *Journal of Food Composition and Analysis*, *22*(3), 257-261.

Pahlawan, M. F. R., & Masithoh, R. E. (2022). Vis-NIR spectroscopy and PLS-DA model for classification of Arabica and Robusta roasted coffee bean. *Advances in Science and Technology*, *115*, 45-52.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International journal of remote sensing*, *26*(1), 217-222.

Panagiotakos, D. B., Pitsavos, C., Chrysohoou, C., Kokkinos, P., Toutouzas, P., & Stefanadis, C. (2003). The J-shaped effect of coffee consumption on the risk of developing acute coronary syndromes: the CARDIO2000 case-control study. *The Journal of Nutrition*, *133*(10), 3228-3232.

Páscoa, R. N., Magalhães, L. M., & Lopes, J. A. (2013). FT-NIR spectroscopy as a tool for valorization of spent coffee grounds: Application to assessment of antioxidant properties. *Food Research International*, *51*(2), 579-586.

Páscoa, R. N., Sarraguça, M. C., Magalhães, L. M., Santos, J. R., Rangel, A. O., & Lopes, J. A. (2015). Use of near-infrared spectroscopy for coffee beans quality assessment. In *Coffee in Health and Disease Prevention* (pp. 933-942). Elsevier.

Pasquini, C. (2003). Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, *14*, 198-219.

Pellegrini, N., Serafini, M., Colombi, B., Del Rio, D., Salvatore, S., Bianchi, M., & Brighenti, F. (2003). Total antioxidant capacity of plant foods, beverages and oils consumed in Italy assessed by three different in vitro assays. *The Journal of Nutrition*, *133*(9), 2812-2819.

Perdoncini, M., Sereia, M. J., Scopel, F. H. P., Formigoni, M., Rigobello, E. S., Beneti, S. C., & Marques, L. (2019). Growth of fungal cells and the production of mycotoxins. *Cell Growth*, *23*.

Pereira, P. V., da Silveira, D. L., Schwan, R. F., de Assis Silva, S., Coelho, J. M., & Bernardes, P. C. (2021). Effect of altitude and terrain aspect on the chemical composition of Coffea canephora cherries and sensory characteristics of the beverage. *Journal of the Science of Food and Agriculture*, *101*(6), 2570-2575.

Perrone, D., Donangelo, C. M., & Farah, A. (2008). Fast simultaneous analysis of caffeine, trigonelline, nicotinic acid and sucrose in coffee by liquid chromatography–mass spectrometry. *Food Chemistry, 110*(4), 1030-1035.

Petracco, M. (2005). The cup. *Espresso coffee: The Science of Quality*, 290-315.

Pham, Y., Reardon-Smith, K., Mushtaq, S., & Cockfield, G. (2019). The impact of climate change and variability on coffee production: a systematic review. *Climatic Change*, *156*, 609-630.

Pizarro, C., Esteban-Díez, I., & González-Sáiz, J.-M. (2007). Mixture resolution according to the percentage of robusta variety in order to detect adulteration in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*, *585*(2), 266-276.

Pizarro, C., Esteban-Díez, I., González-Sáiz, J.-M., & Forina, M. (2007). Use of near-infrared spectroscopy and feature selection techniques for predicting the caffeine content and roasting color in roasted coffees. *Journal of Agricultural and Food Chemistry*, *55*(18), 7477-7488.

Pizarro, C., Esteban-Díez, I., Nistal, A.-J., & González-Sáiz, J.-M. a. (2004). Influence of data pre-processing on the quantitative determination of the ash content and lipids in roasted coffee by near infrared spectroscopy. *Analytica Chimica Acta*, *509*(2), 217-227.

Pohl, P., Stelmach, E., Welna, M., & Szymczycha-Madeja, A. (2013). Determination of the elemental composition of coffee using instrumental methods. *Food Analytical Methods*, *6*, 598-613.

Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science & Technology*, *46*(2), 211-230.

Prashanth, L., Kattapagari, K. K., Chitturi, R. T., Baddam, V. R. R., & Prasad, L. K. (2015). A review on role of essential trace elements in health and disease. *Journal of dr. ntr university of health sciences*, *4*(2), 75-85.

Prevolnik, M., Škrlep, M., Škornjac, D., & Čandek-Potokar, M. (2010). Application of near infrared spectroscopy to predict chemical composition of meat and meat products. *Scientific Journal" Meat Technology"*, *51*(2), 133-142.

Priyadarshini, R., Barik, R. K., & Dubey, H. (2018). Deepfog: fog computing-based deep neural architecture for prediction of stress types, diabetes and hypertension attacks. *Computation*, *6*(4), 62.

Proctor, R. H., McCormick, S. P., Kim, H.-S., Cardoza, R. E., Stanley, A. M., Lindo, L., Kelly, A., Brown, D. W., Lee, T., & Vaughan, M. M. (2018). Evolution of structural diversity of trichothecenes, a family of toxins produced by plant pathogenic and entomopathogenic fungi. *PLoS Pathogens*, *14*(4), e1006946.

Prodolliet, J., Bruelhart, M., Blanc, M. B., Leloup, V., Cherk, G., Donnelly, C. M., & Viani, R. (1995). Adulteration of soluble coffee with coffee husks and parchments. *Journal of AOAC International*, *78*(3), 761-767.

Pugajeva, I., Jaunbergs, J., & Bartkevics, V. (2015). Development of a sensitive method for the determination of acrylamide in coffee using high-performance liquid chromatography coupled to a hybrid quadrupole

Orbitrap mass spectrometer. *Food Additives & Contaminants: Part A*, *32*(2), 170-179.

Qu, F., Ren, D., Hou, J., Zhang, Z., Lu, A., Wang, J., & Xu, H. (2016). The Characteristic Spectral Selection Method Based on Forward and Backward Interval Partial Least Squares. *Guang pu Xue yu Guang pu Fen xi= Guang pu*, *36*(2), 593-598.

Rahim, S. H. A., Ayob, M. K., & Ramli, N. (2011). Fungal contamination of commercial coffee powder. International Seminar on The Current Research Progress in Sciences and Technology (ISSTECH 2011).

Rasch, V. (2003). Cigarette, alcohol, and caffeine consumption: risk factors for spontaneous abortion. *Acta obstetricia et gynecologica Scandinavica*, *82*(2), 182-188.

Rawel, H., & Kulling, S. (2007). Nutritional contribution of coffee, cacao and tea phenolics to human health. *Journal für Verbraucherschutz und Lebensmittelsicherheit*, *2*, 399-406.

Reddy, L., & Bhoola, K. (2010). Ochratoxins—Food contaminants: Impact on human health. *Toxins*, *2*(4), 771-779.

Regulation, C. (2017). Establishing mitigation measures and benchmark levels for the reduction of the presence as acrylamide in food. *Journal of European Union*, *60*, 24-44.

Reh, C., Gerber, A., Prodolliet, J., & Vuataz, G. (2006). Water content determination in green coffee–Method comparison to study specificity and accuracy. *Food Chemistry*, *96*(3), 423-430.

Reis, N., Franca, A. S., & Oliveira, L. S. (2013). Discrimination between roasted coffee, roasted corn and coffee husks by Diffuse Reflectance Infrared Fourier Transform Spectroscopy. *LWT-Food Science and Technology*, *50*(2), 715-722.

Rezende, E. d. F., Borges, J. G., Cirillo, M. Â., Prado, G., Paiva, L. C., & Batista, L. R. (2013). Ochratoxigenic fungi associated with green coffee beans (Coffea arabica L.) in conventional and organic cultivation in Brazil. *Brazilian Journal of Microbiology*, *44*, 377-384.

Ribeiro, J. S., Ferreira, M. M., & Salva, T. (2011). Chemometric models for the quantitative descriptive sensory analysis of Arabica coffee beverages using near infrared spectroscopy. *Talanta*, *83*(5), 1352-1358.

Richelle, M., Tavazzi, I., & Offord, E. (2001). Comparison of the antioxidant activity of commonly consumed polyphenolic beverages (coffee, cocoa, and tea) prepared per cup serving. *Journal of Agricultural and Food Chemistry*, *49*(7), 3438-3442.

Rinnan, Å., Van Den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry*, *28*(10), 1201-1222.

Rodriguez-Saona, L. E., & Allendorf, M. E. (2011). Use of FTIR for rapid authentication and detection of adulteration of food. *Annual Review of Food Science and Technology*, *2*, 467-483.

Rosenberg Jr, C. W. v., Abbate, A., Drake, J., & Mayes, D. M. (2000). A Rugged Near-Infrared Spectrometer for the Real-time Measurement of Grains

During Harvest-An NIR solid-state spectrometer can be used to measure commodity grains in the field in real. *Spectroscopy-Eugene*, *15*(6), 34-39.

Rosipal, R., & Krämer, N. (2006). Overview and recent advances in partial least squares. Subspace, Latent Structure and Feature Selection: Statistical and Optimization Perspectives Workshop, SLSFS 2005, Bohinj, Slovenia, February 23-25, 2005, Revised Selected Papers,

Rubayiza, A. B., & Meurens, M. (2005). Chemical discrimination of arabica and robusta coffees by Fourier transform Raman spectroscopy. *Journal of Agricultural and Food Chemistry*, *53*(12), 4654-4659.

Ruffo Roberto, S., Youssef, K., Hashim, A. F., & Ippolito, A. (2019). Nanomaterials as alternative control means against postharvest diseases in fruit crops. *Nanomaterials*, *9*(12), 1752.

Rutherford, M. A., & Phiri, N. (2006). Pests and diseases of coffee in Eastern Africa: A technical and advisory manual. *Wallingford, UK: CAB International*.

Ryan, D. H. (1996). Medicating the obese patient. *Endocrinology and Metabolism Clinics*, *25*(4), 989-1004.

Ryan, E., Galvin, K., O'connor, T., Maguire, A., & O'brien, N. (2006). Fatty acid profile, tocopherol, squalene and phytosterol content of brazil, pecan, pine, pistachio and cashew nuts. *International journal of food sciences and nutrition*, *57*(3-4), 219-228.

Ryckewaert, M., Gorretta, N., Henriot, F., Marini, F., & Roger, J.-M. (2020). Reduction of repeatability error for analysis of variance-Simultaneous

Component Analysis (REP-ASCA): Application to NIR spectroscopy on coffee sample. *Analytica Chimica Acta*, *1101*, 23-31.

Sacchetti, G., Di Mattia, C., Pittia, P., & Mastrocola, D. (2009). Effect of roasting degree, equivalent thermal effect and coffee type on the radical scavenging activity of coffee brews and their phenolic fraction. *Journal of Food Engineering*, *90*(1), 74-80.

Santos, F. D., Santos, L. P., Cunha, P. H., Borghi, F. T., Romao, W., de Castro, E. V., de Oliveira, E. C., & Filgueiras, P. R. (2021). Discrimination of oils and fuels using a portable NIR spectrometer. *Fuel*, *283*, 118854.

Santos, J. R., Viegas, O., Páscoa, R. N., Ferreira, I. M., Rangel, A. O., & Lopes, J. A. (2016). In-line monitoring of the coffee roasting process with near infrared spectroscopy: Measurement of sucrose and colour. *Food Chemistry*, *208*, 103-110.

Santos, K. M., Moura, M. F., Azevedo, F. G., Lima, K. M., Raimundo Jr, I. M., & Pasquini, C. (2012). Classification of Brazilian coffee using near-infrared spectroscopy and multivariate calibration. *Analytical Letters*, *45*(7), 774-781.

Scalbert, A., Manach, C., Morand, C., Rémésy, C., & Jiménez, L. (2005). Dietary polyphenols and the prevention of diseases. *Critical Reviews in Food Science and Nutrition*, *45*(4), 287-306.

Scarano, D., & Rao, R. (2014). DNA markers for food products authentication. *Diversity*, *6*(3), 579-596.

Schenker, S., Heinemann, C., Huber, M., Pompizzi, R., Perren, R., & Escher, R. (2002). Impact of roasting conditions on the formation of aroma compounds in coffee beans. *Journal of Food Science*, *67*(1), 60-66.

Scholkopf, B., Sung, K.-K., Burges, C. J., Girosi, F., Niyogi, P., Poggio, T., & Vapnik, V. (1997). Comparing support vector machines with Gaussian kernels to radial basis function classifiers. *IEEE transactions on Signal Processing*, *45*(11), 2758-2765.

Schouten, M. A., Tappi, S., & Romani, S. (2020). Acrylamide in coffee: formation and possible mitigation strategies–a review. *Critical Reviews in Food Science and Nutrition*, *60*(22), 3807-3821.

Shan, J., Suzuki, T., Suhandy, D., Ogawa, Y., & Kondo, N. (2014). Chlorogenic acid (CGA) determination in roasted coffee beans by Near Infrared (NIR) spectroscopy. *Engineering in Agriculture, Environment and Food*, *7*(4), 139-142.

Shuping, D., & Eloff, J. N. (2017). The use of plants to protect plants and food against fungal pathogens: A review. *African Journal of Traditional, Complementary and Alternative Medicines*, *14*(4), 120-127.

Siesler, H. W. (2007). Basic principles of near-infrared spectroscopy. In *Handbook of near-infrared analysis* (pp. 25-38). CRC press.

Silva, C. F., Batista, L. R., & Schwan, R. F. (2008). Incidence and distribution of filamentous fungi during fermentation, drying and storage of coffee (Coffea arabica L.) beans. *Brazilian Journal of Microbiology*, *39*, 521-526.

Silva, C. F., Schwan, R. F., Dias, Ë. S., & Wheals, A. E. (2000). Microbial diversity during maturation and natural processing of coffee cherries of Coffea arabica in Brazil. *International Journal of Food Microbiology*, *60*(2-3), 251-260.

Silván, J. M., Morales, F. J., & Saura-Calixto, F. (2010). Conceptual study on maillardized dietary fiber in coffee. *Journal of Agricultural and Food Chemistry*, *58*(23), 12244-12249.

Simmons, K., Kinney, J., Owens, A., Kleier, D., Bloch, K., Argentar, D., Walsh, A., & Vaidyanathan, G. (2008). Comparative study of machine-learning and chemometric tools for analysis of in-vivo high-throughput screening data. *Journal of chemical information and modeling*, *48*(8), 1663-1668.

Slaughter, D., Pelletier, M., & Upadhyaya, S. (2001). Sensing soil moisture using NIR spectroscopy. *Applied engineering in agriculture*, *17*(2), 241.

Smith, A. (1985). A History of Coffee in Coffee: Botany, Biochemistry and Production of Beans and Beverage; Clifford, MN, Willson, KC, Eds. In: Springer-Verlag: New York, NY, US.

Soares, C., Cunha, S., & Fernandes, J. (2006). Determination of acrylamide in coffee and coffee products by GC-MS using an improved SPE clean-up. *Food Additives and Contaminants*, *23*(12), 1276-1282.

Souza, R., & Benassi, M. T. (2012). Discrimination of commercial roasted and ground coffees according to chemical composition. *Journal of the Brazilian Chemical Society*, *23*, 1347-1354.

Spink, J., & Moyer, D. C. (2011). Defining the public health threat of food fraud. *Journal of Food Science, 76*(9), R157-R163.

Stefanov, I., Baeten, V., Abbas, O., Vlaeminck, B., De Baets, B., & Fievez, V. (2013). Evaluation of FT-NIR and ATR-FTIR spectroscopy techniques for determination of minor odd-and branched-chain saturated and trans unsaturated milk fatty acids. *Journal of Agricultural and Food Chemistry, 61*(14), 3403-3413.

Studer, A., Blank, I., & Stadler, R. (2004). Thermal processing contaminants in foodstuffs and potential strategies of control. *Czech Journal of Food Sciences, 22*(I), 1.

Suhandy, D., & Yulia, M. (2017). Peaberry coffee discrimination using UV-visible spectroscopy combined with SIMCA and PLS-DA. *International journal of food properties, 20*(sup1), S331-S339.

Sun, M., Zhang, D., Liu, L., & Wang, Z. (2017). How to predict the sugariness and hardness of melons: A near-infrared hyperspectral imaging method. *Food Chemistry, 218*, 413-421.

Surma, M., Sadowska-Rociek, A., Cieślik, E., & Sznajder-Katarzyńska, K. (2017). Optimization of QuEChERS sample preparation method for acrylamide level determination in coffee and coffee substitutes. *Microchemical Journal, 131*, 98-102.

Svilaas, A., Sakhi, A. K., Andersen, L. F., Svilaas, T., Strom, E. C., Jacobs Jr, D. R., Ose, L., & Blomhoff, R. (2004). Intakes of antioxidants in coffee, wine,

and vegetables are correlated with plasma carotenoids in humans. *The Journal of Nutrition*, *134*(3), 562-567.

Sylvain, P. G. (1958). Ethiopian coffee—its significance to world coffee problems. *Economic botany*, *12*(2), 111-139.

Tang, J., Liao, X., Tong, H., & Gao, J. (2014). GC-MS combined with PLS-DA to discriminate the varieties of XinJiang lavender essential oil. *Computers and Applied Chemistry*, *31*(6), 701-704.

Tavares, K. M., Lima, A. R., Nunes, C. A., Silva, V. A., Mendes, E., Casal, S., & Pereira, R. G. A. (2016). Free tocopherols as chemical markers for Arabica coffee adulteration with maize and coffee by-products. *Food Control*, *70*, 318-324.

Tavares, K., Pereira, R., Nunes, C., Pinheiro, A., Rodarte, M., & Guerreiro, M. (2012). Mid-infrared spectroscopy and chemometrics were used to identify adulteration in roasted and ground coffee by addition of coffee husks. *Química Nova*, *35*, 1164-1168.

Teye, E., & Amuah, C. L. (2022). Rice varietal integrity and adulteration fraud detection by chemometrical analysis of pocket-sized NIR spectra data. *Applied Food Research*, 100218.

Teye, E., Elliott, C., Sam-Amoah, L. K., & Mingle, C. (2019). Rapid and nondestructive fraud detection of palm oil adulteration with Sudan dyes using portable NIR spectroscopic techniques. *Food Additives & Contaminants: Part A*, *36*(11), 1589-1596.

Teye, E., Huang, X., Dai, H., & Chen, Q. (2013). Rapid differentiation of Ghana cocoa beans by FT-NIR spectroscopy coupled with multivariate classification. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *114*, 183-189.

Teye, E., Huang, X.-y., Lei, W., & Dai, H. (2014). Feasibility study on the use of Fourier transform near-infrared spectroscopy together with chemometrics to discriminate and quantify adulteration in cocoa beans. *Food Research International*, *55*, 288-293.

Thapngam, T., Yu, S., & Zhou, W. (2012). DDoS discrimination by linear discriminant analysis (LDA). 2012 International Conference on Computing, Networking and Communications (ICNC),

Thibault, J., Van Breusegem, V., & Chéruy, A. (1990). On-line prediction of fermentation variables using neural networks. *Biotechnology and Bioengineering*, *36*(10), 1041-1048.

Thorburn Burns, D., Tweed, L., & Walker, M. (2017). Ground roast coffee: review of analytical strategies to estimate geographic origin, species authenticity and adulteration by dilution. *Food Analytical Methods*, *10*(7), 2302-2310.

Tian, L., Guo, Y., Zhang, A., & Zhong, H. (2022). Simultaneously Verifying the Original Region of Green and Roasted Coffee Beans by Stable Isotopes and Elements Combined with Random Forest. *Journal of Food Quality*, *2022*.

Ting, D. F., Pui, L. P., & Solihin, M. I. (2020). Feasibility of fraud detection in milk powder using a handheld near-infrared spectroscopy. AIP Conference Proceedings,

Toci, A. T., Farah, A., Pezza, H. R., & Pezza, L. (2016). Coffee adulteration: More than two decades of research. *Critical Reviews in Analytical Chemistry*, *46*(2), 83-92.

Tournas, V., & Katsoudas, E. J. (2008). Microbiological quality of various medicinal herbal teas and coffee substitutes. *Microbiology insights*, *1*, MBI. S943.

Tritscher, A. M. (2004). Human health risk assessment of processing-related compounds in food. *Toxicology Letters*, *149*(1-3), 177-186.

Tugnolo, A., Beghi, R., Giovenzana, V., & Guidetti, R. (2019). Characterization of green, roasted beans, and ground coffee using near infrared spectroscopy: A comparison of two devices. *Journal of Near Infrared Spectroscopy*, *27*(1), 93-104.

Uncu, A. T., & Uncu, A. O. (2018). Plastid trnH-psbA intergenic spacer serves as a PCR-based marker to detect common grain adulterants of coffee (Coffea arabica L.). *Food Control*, *91*, 32-39.

Varga, J., Kocsubé, S., Péteri, Z., Vágvölgyi, C., & Tóth, B. (2010). Chemical, physical and biological approaches to prevent ochratoxin induced toxicoses in humans and animals. *Toxins*, *2*(7), 1718-1750.

Vasconcelos, A. L. S., Franca, A. S., Gloria, M. B. A., & Mendonça, J. C. (2007). A comparative study of chemical attributes and levels of amines in defective green and roasted coffee beans. *Food Chemistry*, *101*(1), 26-32.

Vega, F. E., Posada, F., Aime, M. C., Peterson, S. W., & Rehner, S. A. (2008). Fungal endophytes in green coffee seeds. *Mycosystema*, 27(1), 75-84

Vieira, T., Cunha, S., & Casal, S. (2015). Mycotoxins in Coffee. Coffee in Health and Disease Prevention. In: Elsevier, 225-233, Academic Press.

Virtanen, S., Räsänen, L., Aro, A., Ylönen, K., Lounamaa, R., Akerblom, H., & Tuomilehto, J. (1994). Is children's or parents' coffee or tea consumption associated with the risk for type 1 diabetes mellitus in children? Childhood Diabetes in Finland Study Group. *European Journal of Clinical Nutrition*, *48*(4), 279-285.

Wachamo, H. L. (2017). Review on health benefit and risk of coffee consumption. *Med. Aromat. Plants*, *6*(4), 1-12.

Wang, S., Zhang, Z., Ning, J., Ren, G., Yan, S., & Wan, X. (2013). Back propagation-artificial neural network model for prediction of the quality of tea shoots through selection of relevant near infrared spectral data via synergy interval partial least squares. *Analytical Letters*, *46*(1), 184-195.

Wenzl, T., & Anklam, E. (2007). European Union database of acrylamide levels in food: update and critical review of data collection. *Food Additives and Contaminants*, *24*(sup1), 5-12.

Wilamowski, B. M. (2009). Neural network architectures and learning algorithms. *IEEE Industrial Electronics Magazine*, *3*(4), 56-63.

Williamson, G. (2017). The role of polyphenols in modern nutrition. *Nutrition Bulletin*, *42*(3), 226-235.

Winkler-Moser, J. K., Singh, M., Rennick, K. A., Bakota, E. L., Jham, G., Liu, S. X., & Vaughn, S. F. (2015). Detection of corn adulteration in Brazilian coffee (Coffea arabica) by tocopherol profiling and near-infrared (NIR)

spectroscopy. *Journal of Agricultural and Food Chemistry*, *63*(49), 10662-10668.

Winkler-Moser, J. K., Singh, M., Rennick, K. A., Bakota, E. L., Jham, G., Liu, S. X., & Vaughn, S. F. (2015). Detection of corn adulteration in Brazilian coffee (Coffea arabica) by tocopherol profiling and near-infrared (NIR) spectroscopy. *Journal of Agricultural and Food Chemistry*, *63*(49), 10662-10668.

Wongnaa, C. A., Adams, F., Tempia, N. P., Benjamin, A. B., Charles, B. N. Y., & Sandra, G. (2021). Revamping Ghana's coffee production: smallholder perceptions, viability and constraints. *Thai Journal of Agricultural Science*, *54*(3), 178-197.

Wright Jr, K., Badia, P., Myers, B., & Plenzler, S. (1997). Combination of bright light and caffeine as a countermeasure for impaired alertness and performance during extended sleep deprivation. *Journal of Sleep Research*, *6*(1), 26-35.

Xiaobo, Z., Jiewen, Z., Povey, M. J., Holmes, M., & Hanpin, M. (2010). Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, *667*(1-2), 14-32.

Yan, J., van Stuijvenberg, L., & van Ruth, S. M. (2019). Handheld Near-Infrared Spectroscopy for Distinction of Extra Virgin Olive Oil from Other Olive Oil Grades Substantiated by Compositional Data. *European Journal of Lipid Science and Technology*, *121*(12), 1900031.

Yanagimoto, K., Lee, K.-G., Ochi, H., & Shibamoto, T. (2002). Antioxidative activity of heterocyclic compounds found in coffee volatiles produced by Maillard reaction. *Journal of Agricultural and Food Chemistry*, *50*(19), 5480-5484.

Yergenson, N., & Aston, D. E. (2020). Monitoring coffee roasting cracks and predicting with in situ near-infrared spectroscopy. *Journal of Food Process Engineering*, *43*(2), e13305.

Yu, H., & Kim, S. (2012). SVM Tutorial-Classification, Regression and Ranking. *Handbook of Natural computing*, *1*, 479-506.

Zain, M. Z. M., Shori, A. B., & Baba, A. S. (2017). Composition and health properties of coffee bean. *European Journal of Clinical and Biomedical Sciences*, *3*(5), 97-100.

Yu, H., Wang, Y., & Wang, J. (2009). Identification of tea storage times by linear discrimination analysis and back-propagation neural network techniques based on the eigenvalues of principal components analysis of E-nose sensor signals. *Sensors*, *9*(10), 8073-8082

Zerga, K., & Tsegaye, B. (2020). Coffee Diversity and Conservation in Ethiopia. In *Sustainable Agriculture Reviews 39* (pp. 35-55). Springer.

Zhang, S. J., De Bruyn, F., Pothakos, V., Contreras, G. F., Cai, Z., Moccand, C., Weckx, S., & De Vuyst, L. (2019). Influence of various processing parameters on the microbial community dynamics, metabolomic profiles, and cup quality during wet coffee processing. *Frontiers in Microbiology*, *10*, 2621.

Zhao, S., Qiu, Z., & He, Y. (2021). Transfer learning strategy for plastic pollution detection in soil: Calibration transfer from high-throughput HSI system to NIR sensor. *Chemosphere*, *272*, 129908.

Zhu, M., Long, Y., Chen, Y., Huang, Y., Tang, L., Gan, B., Yu, Q., & Xie, J. (2021). Fast determination of lipid and protein content in green coffee beans from different origins using NIR spectroscopy and chemometrics. *Journal of Food Composition and Analysis*, *102*, 104055.

Zhu, M., Long, Y., Ma, Y., Chen, Y., Yu, Q., Xie, J., Li, B., & Tian, J. (2021). Comparison of chemical and fatty acid composition of green coffee bean (Coffea arabica L.) from different geographical origins. *Lwt*, *140*, 110802.

Zinedine, A., Brera, C., Elakhdari, S., Catano, C., Debegnach, F., Angelini, S., De Santis, B., Faid, M., Benlemlih, M., & Minardi, V. (2006). Natural occurrence of mycotoxins in cereals and spices commercialized in Morocco. *Food Control*, *17*(11), 868-874.

APPENDICES

APPENDIX A- ANOVA results for the proximate results of pure coffee, coffee husk and adulterated coffee

Sample	% Moisture	% Ash	% Protein	% Lipids	% Fibre	% CHO
CRD	2.53±0.32^{ab}	3.62±0.28^{d}	16.18 ± 0.10^{c}	5.45±0.29^{de}	5.75±0.08^{b}	68.97±0.66^{a}
ADH	1.38±0.33^{d}	4.49±0.35^{c}	21.32 ± 0.19^{ab}	7.38±0.16^{a}	5.27±0.13^{cd}	61.55±0.76^{d}
BDH	1.65±0.22^{cd}	6.23±0.20^{a}	21.65 ± 0.52^{a}	7.36± 0.23^{a}	5.08±0.07^{d}	59.62±0.87^{e}
CDH	1.61±0.16^{cd}	5.04±0.03^{bc}	21.08 ± 0.21^{ab}	6.46±0.21^{b}	5.45±0.18b^{cd}	61.94±0.25^{d}
DDH	1.85±0.09b^{cd}	4.49±0.32^{c}	20.72 ± 0.25^{b}	5.78±0.16c^{d}	5.55±0.12b^{c}	63.43±0.09^{c}
EDH	2.04±0.16b^{cd}	5.38±0.05b	21.81 ± 0.36^{a}	5.41±0.27d^{e}	5.44±0.20b^{cd}	61.93±0.38^{d}
FDH	2.28±0.25^{bc}	4.62±0.36^{c}	20.63 ± 0.13^{b}	5.18±0.06e	5.45±0.15b^{cd}	64.10±0.24^{c}
CHP	3.23±0.58^{a}	6.26±0.19^{a}	14.59 ± 0.19^{d}	6.22±0.10^{bc}	6.18±0.06^{a}	66.72±0.37^{b}

Values represent means ± SD of at least three replicates. Means in the same column with the different superscripts are significantly different (p>0.05). CDR- Pure coffee, ADH- 95/5H, BDH- 90/10H, CDH- 85/15H, DDH- 80/20H, EDH- 75/25H, FDH- 70/30H

APPENDIX B- ANOVA results for Polyphenols and Antioxidants of pure coffee, coffee husk and adulterated coffee

Sample	Polyphenols (mg/kg)	Antioxidant (mg/kg)
CRD	4373.30±0.65^{a}	4416.80±0.18^{a}
ADH	3954.90±0.69^{b}	4146.21±0.10^{b}
BDH	3677.79±0.10^{c}	4146.40±0.16^{b}
CDH	3521.40±0.75^{d}	4124.60±0.93^{b}
DDH	3159.00±0.70^{e}	3669.30±0.12^{c}
EDH	3294.50±0.32^{f}	3667.10±0.42^{c}
FDH	3058.50±0.64^{g}	3815.00±0.17^{c}
CHP	354.30±0.29^{h}	739.67±0.74^{d}

Values represent means ± SD of at least three replicates. Means in the same column with the different superscripts are significantly different (p>0.05). CDR- Pure coffee, ADH- 95/5H, BDH- 90/10H, CDH- 85/15H, DDH- 80/20H, EDH- 75/25H, FDH- 70/30H

APPENDIX C

Appendix C 1: Application for Ethical Clearance

APPLICATION LETTER

Vida Gyimah Boadu
School of Agricultural and Natural Resources
Department of Agricultural Engineering
23 June 2023

The Chairperson
Institutional Review Board
University of Cape Coast
Cape Coast

Dear Chairperson,

APPLICATION FOR ETHICAL APPROVAL

Protocol Name:

DEVELOPING NOVEL ONSITE DETECTION TECHNOLOGY BY USING CHEMOMETRICAL ANALYSIS OF HAND-HELD NEAR-INFRARED SENSOR TECHNIQUE FOR COFFEE QUALITY

I wish to submit to you the above-named protocol and essential documents for approval by your board. I look forward to receiving any comments that you may have with the above.

Thank you for your co-operation.

Yours sincerely,

Vida Gyimah Boadu
(Investigator name)

Enclosed:

1. Informed Consent Form
2. Protocol Submission checklist
3. Research Instrument
4. CV of PIs and supervisor

APPENDIX C2: Introduction Letter for Ethical Review

UNIVERSITY OF CAPE COAST
COLLEGE OF AGRICULTURE AND NATURAL SCIENCES
SCHOOL OF AGRICULTURE
DEPARTMENT OF AGRICULTURAL ENGINEERING

Telephone: +233-03321-32709
Head: +233-0506373761
Email: agric.engineering@ucc.edu.gh
Website: www.agric.eng.ucc.edu.gh

UNIVERSITY POST OFFICE
CAPE COAST, GHANA

Our Ref.: SA/DAE/20'
Your Ref.:

23rd June, 2023

The Chairman
Institutional Review Board
University of Cape Coast

Dear Sir,

INTRODUCTORY LETTER FOR ETHICAL CLEARANCE

I have the pleasure of introducing Vida Gyimah Boadu (Student ID number: **AG/DHP/19/0002**), a PhD Student of this department researching on the topic **"Developing Novel Onsite Detection Technology by using Chemometrical Analysis of Hand-Held Near Infrared(NIR) Sensor Technique for Coffee Quality"**.

As part of her study requirements, she is applying for ethical clearance to enable her collect data. I would be grateful if you could give her the necessary assistance.

Counting on your cooperation.

Thank you.

Yours faithfully,

Prof. Ernest Ekow Abano
(Head of Department)

APPENDIX C3: Ethical Clearance

UNIVERSITY OF CAPE COAST

INSTITUTIONAL REVIEW BOARD SECRETARIAT

TEL: [redacted]
E-MAIL: [redacted]
OUR REF: IRB/C3/Vol.1/0399
YOUR REF:
OMB NO: 0990-0279
IORG #: IORG0011497

31ST OCTOBER, 2023

Ms Vida Gyimah Boadu,
Department of Agricultural Engineering
University of Cape Coast

Dear Ms Boadu,

ETHICAL CLEARANCE – ID (UCCIRB/CANS/2023/16)

The University of Cape Coast Institutional Review Board (UCCIRB) has granted Provisional Approval for the implementation of your research **Developing Novel Onsite Detection Technology by using Chemometrical Analysis of Hand-Held Near-Infrared Sensor Technique for Coffee Quality.** This approval is valid from **31st October, 2023** to **30th October, 2024.** You may apply for an extension of ethical approval if the study lasts for more than 12 months.

Please note that any modification to the project must first receive renewal clearance from the UCCIRB before its implementation. You are required to submit a periodic review of the protocol to the Board and a final full review to the UCCIRB on completion of the research. The UCCIRB may observe or cause to be observed procedures and records of the research during and after implementation.

You are also required to report all serious adverse events related to this study to the UCCIRB within seven days verbally and fourteen days in writing.

Always quote the protocol identification number in all future correspondence with us in relation to this protocol.

Yours faithfully

Kofi F. Amuquandoh
Ag. Administrator