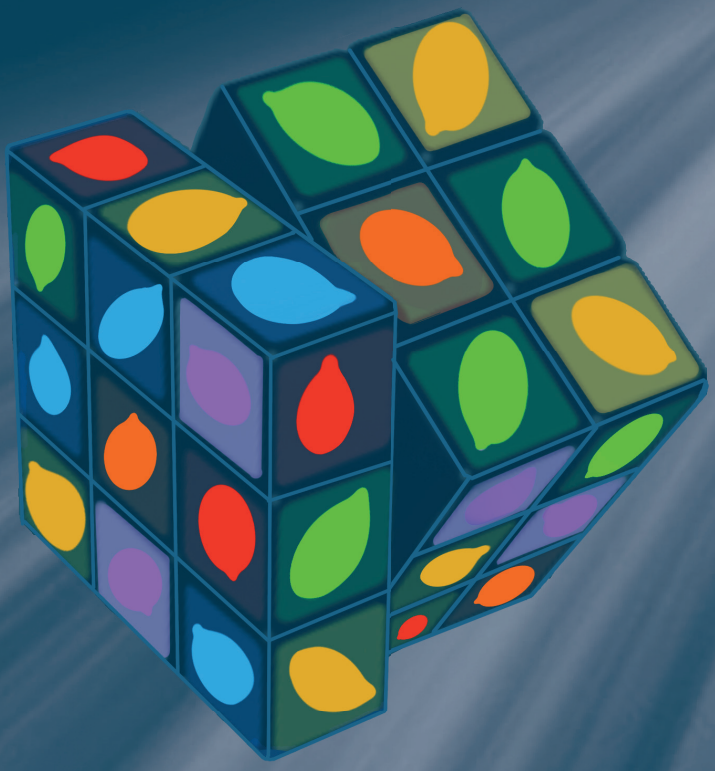


NOVEL SPECTROSCOPIC APPROACHES FOR THE CHARACTERISATION OF QUALITY-AND IDENTITY-RELATED KEY FEATURES OF PEANUTS AND PEANUT BUTTERS

HONGWEI YU

Propositions

1. The portable near infrared spectrometer has great applications in the peanut industry chain.
   (this thesis)

2. Classification based on the peanut processing utility has the most practical value compared to other classification approaches.
   (this thesis)

3. Reference samples and reference measurements play critical roles in the development of screening methods.

4. It is the best of times and the worst of times for data-driven technology as well.

5. By learning to observe the world we are taught answers to scientific questions.

6. Every mountain to be conquered in life starts with climbing the first step.

7. An individual can cycle fast, but a group can cycle far.

Propositions belonging to the thesis, entitled

Novel spectroscopic approaches for the characterisation of quality- and identity-related key features of peanuts and peanut butters

Hongwei Yu
Wageningen, 10 June 2022

Novel spectroscopic approaches for the characterisation of quality- and identity-related key features of peanuts and peanut butters

Hongwei Yu

Thesis committee

Promotors
Prof. Dr S.M. van Ruth
Special Professor, Food Authenticity and Integrity
Wageningen University & Research

Prof. Dr Q. Wang
Professor, Institute of Food Science and Technology
Chinese Academy of Agricultural Sciences, China

Other members
Prof. Dr H.A. Schols, Wageningen University & Research
Dr J.J. Jansen, Radboud University, Nijmegen
Dr V. Baeten, CRA-W Wallonie Recherche, Belgium
Dr D. Wu, Queen's University Belfast, UK

This research was conducted under the auspices of the Graduate School VLAG (Advanced studies in Food technology, Agrobiotechnology, Nutrition and Health Sciences)

Novel spectroscopic approaches for the characterisation of quality- and identity-related key features of peanuts and peanut butters

Hongwei Yu

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Friday 10 June 2022
at 11 a.m. in the Omnia Auditorium.

Hongwei Yu

Novel spectroscopic approaches for the characterisation of quality- and identity-related key features of peanuts and peanut butters,
160 Pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2022)
With references, with summary in English

ISBN: 978-94-6447-204-2
DOI: https://doi.org/10.18174/568518

Table of contents

CHAPTER 1

General introduction

1.1. PEANUT

1.1.1. Peanut resources

Peanut (*Arachis hypogaea L.*), also known as groundnut, originated in Latin America (Hammons et al., 2016). The *Arachis* is a larger genus that has over 70 species, and the only widely grown *Arachis* specie is the cultispecies peanut, whereas the rest are all wild relatives that are either seldom grown or utilised in different ways (Bertioli et al., 2011; Wang, 2018). There are over 40,000 peanut germplasm resources across the world. Among all institutions and countries with abundant peanut germplasm resources, the International Crop Research Institute for the Semi-Arid Tropics (ICRISAT) has the most germplasm resources (15,342), followed by the United States of America (the USA) (8,719), China (7,490), Argentina (2,200), Indonesia (1,730), etc. (Wang, 2018). As a major commercial crop, peanuts commonly prevail in the tropics, the subtropics, and warm temperate zones of the earth, shown in **Figure 1.1**.

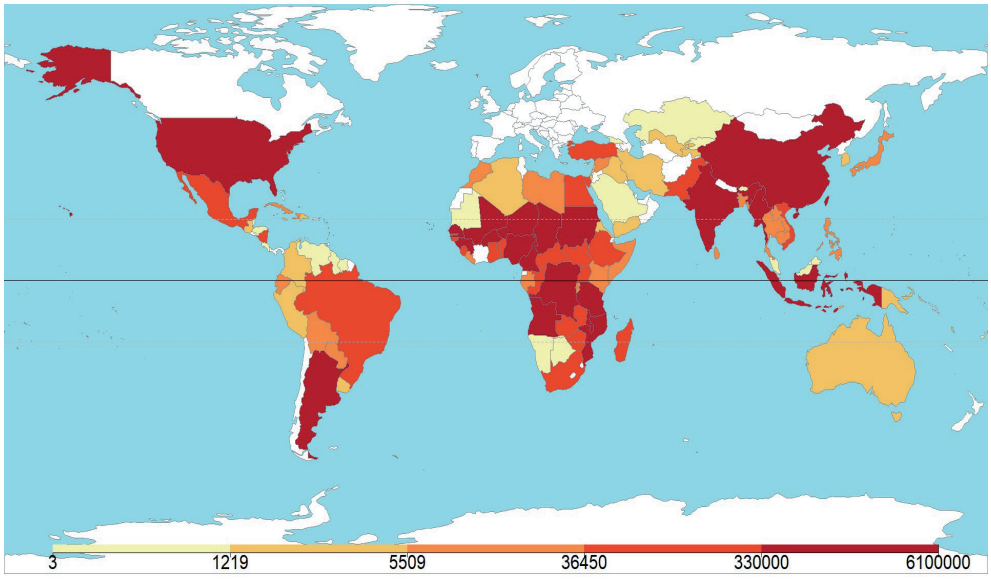


Figure 1.1. The main planting areas and countries of peanuts (adapted from Faostat (2020)). The colour from red to orange refers to high and low area harvested. The solid line is the Equator.

On the report of the Food and Agriculture Organization (FAO) statistics (Faostat, 2020), the average world peanut production from 2010 to 2019 was 45.8 million tonnes and the average area harvested was 27 million ha. Among all countries, China has the largest peanut production with 16.7 million tonnes, accounting for nearly 37% of worldwide production, followed by India (7.4 million tonnes), Nigeria (3.7 million tonnes), the USA (2.4 million tonnes), and Sudan (1.9 million tonnes). India has the most prominent area

harvested (5.1 million ha), followed by China (4.6 million ha), Nigeria (3.1 million ha), Sudan (2.3 million ha), and Myanmar (1.0 million ha).

1.1.2. Peanut characteristics

Peanuts are not only consumed as raw materials or after simple pre-processing such as roasting and boiling, but they are also served as finished goods such as peanut oil, peanut butter, nut bars, and confections (Wang, 2016). Peanuts are appreciated worldwide because they are affordable, flavourful, and nutritious food. The average values of major (fat, protein, and carbohydrates) and minor (amino acids, minerals, and vitamins) constituents in the primary market peanuts are shown in **Figure 1.2**. These data are collected from the Database of Chinese Peanut Processing Characteristics and Special Varieties (Wang, 2020) and the United States Department of Agriculture (USDA) National Nutrient Database for Standard Reference (Haytowitz, 2021). These constituents' values as the analytical signatures of peanuts vary substantially with variety, growing environment, storage condition, and maturity.

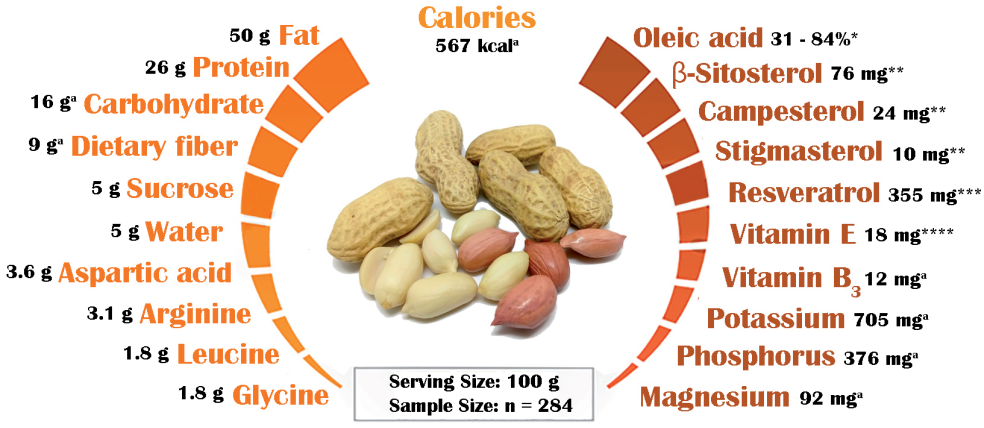


Figure 1.2. The average values of major and minor constituents in peanuts (adapted from Wang (2020) and Haytowitz (2021), [a] means that data come from Haytowitz). * the proportion in fat; ** n = 45; *** the sample is peanut root, n = 19; **** n = 199.

Fat is the predominant macro component in peanuts, accounting for nearly 50% (**Figure 1.2**). Monitoring the fat content before manufacturing is essential for peanut-based commodities such as peanut oil and peanut butter since variations have an impact on the quality traits of the final products. In consideration of economic interest, oil producers anticipate high fat content varieties from breeders. Most fatty acids (FAs) in peanuts are generally presented as triglycerides, which are three FAs esterified to a glycerine backbone. The primary FAs mentioned among all cultivated peanut varieties are palmitic acid (C16:0), oleic acid (C18:1 n9c), and linoleic acid (C18:2 n6c), and these typically make up

above 90% of the total FAs (Bera et al., 2019). FAs are the key compositional factors in the peanut trade since they have important contributions to human health and production stability (Gong et al., 2018; Zhao et al., 2019).

Peanuts have the highest protein level of frequently consumed nuts (Venkatachalam & Sathe, 2006), with a reported value of about 26%, as shown in **Figure 1.2**. Given our planet's rapidly growing population, the world's protein resources are currently under severe strain. Peanut protein, as a critical plant protein, can be studied from manifold quality aspects, involving human and animal nutrition and functional characteristics. In terms of nutrition, peanut protein is a fundamental and essential component for human development (food) and animal vigour (feed) with the biological value (59), the net protein utilisation rate (51%), and the pure digestibility (90%) (Wang, 2016). Regarding functional characteristics, peanut protein has great gelation or solubility based on the constitution of protein, which has already been applied to sausage (gelation) and beverage (solubility) (Wang, 2018). Amino acids are the essential building blocks of proteins. Aspartic acid (~ 3.6%) and arginine (~ 3.1%) are the most common amino acids in peanuts (Wang et al., 2013). In comparison with typical tree nuts, peanuts are inherently rich in the above-mentioned acidic amino acids, as well as hydrophobic amino acids, such as leucine, glycine, and alanine (Davis & Dean, 2016; Wang, 2016).

The total carbohydrate types contain sugars, starch, dietary fibre, long polysaccharides, etc. The total carbohydrate content of peanuts is regularly estimated to be about 16% (**Figure 1.2**). Myo-inositol, glucose, fructose, sucrose, raffinose, and stachyose are the sugars detected, with sucrose accounting for almost 90% of the sugars (Davis et al., 2016). Sugars, especially sucrose, play a vital role in the nonenzymatic processes that give roasted peanut products their distinctive colour and taste. Roasted peanut products have a darker appearance under the same protocols since those peanut varieties have higher sugar content (Davis et al., 2016). Moreover, higher sugar content inclines peanuts to generate a fruity fermented flavour after roasting. Because sugar concentration in peanuts is rather important for roast quality, manufacturers and processors keep a close eye on sugar content to promptly regulate roasting settings to redress these variations and improve the quality traits of final commodities.

Peanuts are regarded as one of the most important mineral resources for nut consumers. Typically, the macrominerals are minerals that exist in concentrations greater than 0.01% of the bodyweight, including calcium, magnesium, phosphorus, potassium, sodium, etc., while the rest are the microminerals including copper, iodine, iron, manganese, zinc, etc. The mineral content in peanuts is only 2 - 3%. The potassium (705 mg), phosphorus (376 mg), and magnesium (92 mg) contents in peanuts are high, whereas calcium, iodine, and iron contents are low. Peanuts, like most oilseeds, are a poor source

of the oil-soluble vitamins A, D, and K, but they have abundant vitamin E (18 mg). Therefore, peanuts and peanut butters are excellent resources of vitamin E, providing more than 10% of the daily dietary recommended value with one serving (Haytowitz, 2021). In addition, peanuts are also good resources of B vitamins, containing high contents of B_3 (12 mg) and B_6 (~ 0.5 mg).

1.1.3. Peanut classifications

Peanut classifications can maximise the benefits of peanut planting, processing, and trading. There are different standards for peanut classifications. For instance, peanuts can be divided into spring, summer, and autumn peanuts according to different sowing dates (de Oliveira Aparecido et al., 2021). Peanuts can also be divided based on variety and the length of the growth period, specifically early-maturing, middle-maturing, and late-maturing (Culbreath et al., 1999). The above-mentioned types of classification methods are based on the agronomic characteristics of peanuts. However, some more commercial standards prevail in the markets.

There are four classes of peanuts commonly used in the USA according to their size and appearance: Runner, Virginia, Spanish, and Valencia (Chukwumah et al., 2012). Runner peanuts have acquired broad popularity because of their appealing kernel size range. They account for about half of the peanut production in the USA and most of them are utilised to produce peanut butter (Shin et al., 2010). Virginia peanuts have the biggest kernels and are mainly used for processing roasted peanuts. Smaller kernels with reddish-brown skin are Spanish peanuts, which are mostly utilised in peanut candy. There are typically three or more small kernels in the Valencia peanut pods. Valencia peanuts are sweet and are typically roasted and served in the shell (Archer, 2016; Toomer, 2018).

Furthermore, another standard is to classify peanuts based on their compositions. Some peanut varieties with a high content of some components such as fat, protein, and sucrose have high commercial values and great success in the marketplace. Currently, high oleic acid peanuts (HOP) get the most attention because they cannot only enhance the shelf life of the final products such as peanut butter (Gong et al., 2018), but they can also boost nutritional values (Zhao et al., 2019). HOP have higher oleic acid (80%), lower linoleic acid (3%), and lower palmitic acid (5%) than regular peanuts (RP) but there is no noticeable difference in appearance and size (Talcott et al., 2005). As a result, there are potential risks of using RP indiscriminately as HOP in the markets.

Peanut classification based on processing purposes is becoming mainstream. Main peanut finished goods, involving peanut protein and peanut oil, have distinct requirements for the raw materials. In other words, the characteristics of the products prepared

from different peanut varieties are diverse. The principles of these classification standards are derived from the relationships between peanut and its products. For example, it has been shown that peanut protein with a high gel property level has a positive relationship with the arachin content and the arachin/conarachin protein ratio, as different peanut varieties could be classified into three distinct groups to manufacture peanut protein with different gel property levels (Wang et al., 2014). Furthermore, the induction time of peanut oil has been significantly correlated with unsaturated FAs and the oleic acid/linoleic acid ratio. Accordingly, different peanut varieties were grouped into three distinct types to produce peanut oil with different induction time levels (Zhang, 2012). The processors can further select the specific peanut varieties to produce high-quality products. Therefore, peanut classification based on processing purposes has the potential to better ensure the quality traits of peanut products and improve their market values.

1.2. PEANUT BUTTER

1.2.1. Peanut butter manufacture and consumption

Peanut butter is a dense mixture of solid particles suspended in a consecutive oil phase (Norazatul Hanim et al., 2016). Being rich in fat, protein, vitamins, minerals, and other nutrients, peanut butter is one of the most popular and significant plant-based spreads in the world with a unique flavour and texture (Gong et al., 2018), as well as health advantages such as lowering the risk of gastric non-cardia adenocarcinoma among American seniors (Hashemian et al., 2017). The manufacturing process of peanut butter is standardised. The primary processing starts with the pre-cleaning and shelling of peanuts. The kernels are then transported to the oven for roasting. The roasting parameters, such as temperature and time, should be adequately managed since it affects the formation of aroma and colour of the finished peanut butter products. After roasting, the kernels are typically browned and the skins are loosened. It is necessary to remove the skins by gentle brushing. Discoloured and other rejected kernels are removed before the peanuts are milled. The milling process is implemented to diminish the dimensions of kernels to produce peanut butter. The milling process can be repeated to obtain the desired final texture (Gorrepati et al., 2015; Shakerardekani et al., 2013).

As stated in a business report (SRD, 2020), the annual worldwide production of peanut butter is nearly 150,000 tonnes. The USA is the world leader in peanut butter production and consumption (50,000 tonnes). In terms of the markets in Europe, about 48,000 tonnes of peanut butter are consumed per year. The Centre for the Promotion of Imports (CBI) reported that the Netherlands is Europe's largest processor of peanut and largest manufacturer of peanut butter. Peanuts for processing are mainly imported from Argen-

tina, America, and China (CBI, 2020). In China, the peanut butter processing factories are primarily located in Shandong province where Luhua and Huayu peanut varieties are mostly planted. The annual production of peanut butter in China is about 10,000 tonnes, with imports and domestic markets evenly split. Compared with the USA and Europe, the consumption of peanut butter in China has great potential for growth.

1.2.2. Peanut butter characteristics

The characteristics of peanut butters are important factors used to determine their market values. These characteristics can be divided into three categories: fundamental safety traits (e.g. *Escherichia coli*, Salmonella, acid value, etc.), compositional traits (e.g. water, fat, protein, etc.), and sensory traits (e.g. colour, texture, volatile compounds, etc.) (Gong et al., 2018; NY/T958, 2006). Some countries have already stipulated the various criteria of safety and compositional qualities for peanut butter (IS9037, 1979; NY/T958, 2006). According to the USDA standards (USDA, 1972), peanut butters can be graded into different types according to evaluation traits by sensory analysis. Sensory analysis offers the direct connection between marketing and technology, allowing for a more entire picture of customer perceptions and preferences. However, on one hand, there are some drawbacks to impede the applications of sensory analysis such as inconsistent evaluation criteria and the correspondingly laborious progress (Fan et al., 2016). On the other hand, sensory analysis cannot be used in some circumstances to generate meaningful data (Kilcast, 2013). Therefore, some instrumental methods have been used to scientifically describe the quality traits of peanut butters (Shakerardekani et al., 2013).

Peanut butter's most obvious feature is colour, which is formed by browning reactions and caramelisation during roasting. For instance, Ha et al. (2013) reported colour values of 56 (L*, the lightness value), 10 (a*, the green (negative) - red (positive) opponent colours), and 26 (b*, the blue (negative)-yellow (positive) opponent colours), respectively. Food texture is one of the physical features that considerably influences consumer attraction, purchase decision, and eventual consumption. Texture properties are normally characterised by firmness which is a critical trait of semi-solid food texture. And and Resurreccion (2002) reported a firmness value of 93.3 g in natural peanut butter. Similarly, peanut butter's flow and deformation under tension and stress are known as rheological properties. Rheological data are required to determine the practicality of ingredients during product development, immediate or finished goods management and process engineering calculation for instruments such as pumps and mixers (Shakerardekani et al., 2013). Peanut butter displays complex characteristics of colloidal particle suspensions, and natural peanut butter is a Newtonian fluid with a yield stress of 24 Pa (Citerne et al., 2001). Furthermore, particle size not only affect the mouthfeel of peanut butter and particle-particle interactions also has profound influences on the rheological properties (Tanti et al., 2016). According to a

previous study (Liedl Jr & Rowe, 2007), satisfactory sensory and textural characteristics of nut spread could be manufactured if 90% of particles are smaller than 40 μm. The attractive aromas of peanut butter are generated by means of reactions such as the Maillard reactions during the roasting and milling phases. The main volatile compounds in peanut butter are composed of pyrazines, aldehydes, furans, pyrroles, and ketones (Lou et al., 2009). Among them, pyrazines, including 2,5-dimethylpyrazine and 2,3,5-trimethylpyrazine, are the most important volatile compounds (Chetschik et al., 2010). The above instrumental analysis, as a complement to sensory evaluation, objectively and quantitatively describes the characteristics of peanut butters.

1.2.3. Peanut butter classifications

Three sorts of peanut butters prevail in the markets, including natural peanut butter, creamy or smooth peanut butter, and chunky peanut butter. Natural peanut butter is composed of 100% peanuts without any additional ingredients. Conversely, the formulations of the two others are based on natural peanut butter with added sugar and hydrogenated vegetable oil which guarantee that the consistency of peanut butter is unaffected by time or temperature. Chunky peanut butter is furtherly based on smooth peanut butter with peanut particles. Furthermore, according to the USDA standards, peanut butters could be rated into classes A and B (USDA, 1972). Specifically, peanut butters could be ascertained on the basis of colour, consistency, absence of defects, as well as flavour and aroma, in addition to the other requirements of the respective grade. For instance, the colour for class A represents a rich colour typical of peanut butter manufactured from fittingly roasted peanuts. The flavour and aroma for class A indicate a flavour and aroma: typical of freshly roasted and freshly ground peanuts and free from distasteful flavours and obnoxious odours of any kind. The above-mentioned criteria are mainly used to classify/grade peanut butter based on its composition or sensory quality, but some other evaluations have been used for the classification of other agricultural products in recent years. For instance, red wine could be classified based on phenolic profiles, colour intensity, and potassium content due to grape varieties (Heras-Roger et al., 2016); soymilk attributes could be classified based on the quality traits of the glycinin (11S)/β-conglycinin (7S) protein ratio, soluble solids, and fat content (Ma et al., 2015); chapatti attributes could be assessed in accordance with the physicochemical, rheological, and sensory traits for selecting wheat varieties (Kundu et al., 2017). Therefore, the opportunity exists for peanut butters to be evaluated and classified based on the corresponding quality traits resulting from the use of raw materials.

1.3. THE RELATIONSHIP BETWEEN PEANUT AND PEANUT BUTTER

Agricultural raw materials, known as "the first workshop" of processing, form the basis of the corresponding processing products. Different varieties have various impacts on the quality traits of products under the same processing conditions due to the huge variances in the physical characteristics and chemical compositions of these varieties. Just like the influences of wheat varieties on chapattis (Kundu et al., 2017), olive varieties on extra virgin olive oils (Deiana et al., 2019), grape varieties on sparkling wines (Pérez-Magariño et al., 2015), and apple varieties on purees (Lan et al., 2020), the overall characteristics of peanut butters are related to the quality traits of peanut varieties. Generally, most studies merely considered one or two quality traits of peanut butters derived from a limited number of peanut varieties. For instance, Mohd Rozalli et al. (2015) specifically examined the diameters of particles and the dynamic rheological characteristics of peanut butters, indicating that Indian peanut butter had a greater dynamic modulus than Chinese peanut butter at all respective grinding times. Norazatul Hanim et al. (2016) discovered that peanut butters made from two distinct varieties had various particle size distributions. The polydispersity values of peanut butters manufactured from Spanish varieties were certainly greater than Virginia varieties, indicating that peanut butters derived from Spanish varieties had wider particle size distributions. Lou et al. (2009) investigated the volatile compounds of two distinct peanut butters as well. The results showed that 62 volatile compounds were identified in one sample with higher pyrazine contents, while the other sample only contained 42 compounds. Some scientists (Dhamsaniya et al., 2012; Gong et al., 2018) have already studied the impacts of various chemical contents of peanuts on the textural and sensory qualities of peanut butters, as well as their stability during storage. Dhamsaniya et al. (2012) reported that Somnath composed of 49% fat and 21% protein yielded better firmness and flavour, and was the recommended variety for peanut butter manufacture. Gong et al. (2018) stated that peanut butter manufactured with Kainong 1715 (high oleic acid variety) had a significantly longer shelf-life than other peanut varieties. Even though certain relationships between them have been discovered, the picture is still incomplete.

1.4. ANALYSIS METHODS OF QUALITIES AND CLASSIFICATIONS

1.4.1. Lab-based confirmatory techniques

To analyse the quality traits of peanuts and peanut butters, several lab-based confirmatory methods have been established, with some of them becoming standards. For instance, peanuts are analysed for fat by the Soxhlet extraction (GB5009.6, 2016), crude protein by the Kjeldahl procedure (GB5009.5, 2016) with a conversion coefficient of 5.46, amino acids by high performance liquid chromatography (HPLC) (GB5009.124, 2016), sucrose content by HPLC (GB5009.8, 2016), and FAs profiles by gas chromatography (GC)

(GB5009.168, 2016). Correspondingly, peanut butters are analysed for volatile compounds by GS-MS (Lou et al., 2009), colour by the colour spectrophotometer (Shakerardekani et al., 2013), texture qualities by the texture analyser (Mohd Rozalli et al., 2016), rheology by the rheometer (Norazatul Hanim et al., 2016), and particle size by the technique of laser diffraction (Mohd Rozalli et al., 2015). These methods provide reliable and detailed data on the quality and classification of peanuts and peanut butters. Despite this, the above methods are generally relatively time-consuming and costly. In addition, these methods are literally conducted by trained personnel, have various complicated operations and can relatively easily be error-prone and manipulated during analysis. With the increasing requirements for detection of the quality and identity of peanuts and their products, a range of non-destructive and high throughput screening techniques combined with chemometrics have been widely applied in the peanut industry as efficient, simple, and sensitive analytical methods for quality control and management.

1.4.2. Screening techniques

Vibrational spectroscopy, including near-infrared spectroscopy (NIRS), hyperspectral imaging (HSI), and Raman spectroscopy, has been extensively utilised for the evaluation of the quality and identity of peanuts and their products (**Table 1.1)**. NIRS has broad applications in the quality evaluation of peanuts, such as for the determination of water (Govindarajan et al., 2009b), fat (Fox & Cruickshank, 2005), protein and amino acids (Wang et al., 2013), FAs contents of peanuts (Wu et al., 2009), as well as acid value of peanut oil (Rao et al., 2009). NIRS has been employed to qualitatively and quantitatively examine adulterants in pure peanut oil as well (Castro et al., 2021; Zeng et al., 2016). For single kernel peanuts, Tillman et al. (2006) reported that NIRS can be used to accurately predict FAs contents. Compared with NIRS, HSI can provide additional spatial information, offering new insights into the distribution of components in peanuts like fat (Sun et al., 2020) and protein (Cheng et al., 2018). More importantly, HSI has a great advantage for fungi-contaminant and mouldy peanut classification (Qiao et al., 2017; Yuan et al., 2020). Raman spectroscopy is used for peanut and peanut oil identification as well (Farber et al., 2020; Jin et al., 2021).

The above-mentioned screening methods are based on benchtop applications. However, the miniaturisation of analytical instruments has allowed for on-site testing over the past decade. These techniques can provide user-friendly and speedy screening for the samples, prior to further lab-based confirmatory tests. Portable NIRS spectrometers have been applied for the chemical analysis of peanuts (Bilal et al., 2020) and acid values of peanut oil (Yang et al., 2017). Although the sensitivity is lower than benchtop counterparts in conjunction with routinely updated chemometrics models, such an approach might assist to assure the quality and identity of peanuts and products derived from them. Meanwhile, it is incredibly valuable for breeders and other stakeholders to detect the quality traits of

Table 1.1. Examples of screening techniques for peanuts and their products.

Method	Chemometrics	Applications	Reference
NIRS (Benchtop)	PLS/PLS-DA	Quantitative and qualitative determination of acid value in peanut oil	(Rao et al., 2009)
NIRS (Benchtop)	PLS	Quantitative determination of water in peanuts	(Govindarajan et al., 2009)
NIRS (Benchtop)	PLS	Quantitative determination of protein and amino acid in peanuts	(Wang et al., 2013)
NIRS (Benchtop)	PLS	Quantitative determination of essential minerals in peanuts	(Phan-Thien et al., 2011)
NIRS (Benchtop)	PLS	Quantitative determination of FAs in peanut oil	(Wu et al., 2009)
NIRS (Benchtop)	PLS	Quantitative determination of FAs in single peanut seeds	(Tillman et al., 2006)
NIRS (Benchtop)	SVM	Authentication in peanut oil	(Zeng et al., 2016)
NIRS (Portable)	PLS/GA-PLS	Quantitative determination of acid value in peanut oil	(Yang et al., 2017)
NIRS (Portable)	PLS/GA-PLS/ Si-PLS	Quantitative determination of fat, protein, fibre, carbohydrate, water, ash in peanuts	(Bilal et al., 2020)
HSI	PLS	Quantitative determination of fat in peanuts	(Yu et al., 2016)
HSI	PLS	Quantitative determination of protein in peanuts	(Yu et al., 2017)
HSI	SVM	Identification of fungi contaminant in peanuts	(Qiao et al., 2017)
HSI	PCA	Identification of fungi contaminant in peanuts	(Jiang et al., 2016)
Raman	PLS-DA	Identification of adulterated peanut oil	(Huang et al., 2016)

FAs, Fatty acids; GA-PLS, Genetic algorithm Partial least squares; HSI, Hyperspectral imaging; NIRS, Near-infrared spectroscopy; PCA, Principal component analysis; PLS, Partial least squares; PLS-DA, Partial least squares discrimination analysis; Si-PLS, Synergy interval partial least squares; SVM, Support vector machine.

single peanut kernels to further choose the required varieties for large-scale cultivation. However, NIRS, as the bulk sample spectrometer, needs roughly 250 g of peanut kernels (Agelet & Hurburgh, 2014). The demands of rapid analysis of single peanut kernels are enormous.

1.4.3. Data analysis

Many measurements discussed previously lead to huge amounts of datasets from test samples containing analytical signatures. The development of reference datasets that provide complete and standardised information based on representative samples is essential for establishing reliable screening methods. Due to the multiple chemical or physical properties included, it is challenging to interpret data directly. Diverse statistical methods have been widely implemented to explore the ability of a large dataset to determine the components quantitatively and identify varieties (Alewijn et al., 2016; Granato, Putnik, et al., 2018). First of all, univariate analysis, including descriptive analysis (e.g. calculation of range) and statistical inference (e.g. analysis of variances), are used to explore and

infer the data (Yang et al., 2020). And then, multivariate analysis is sequentially applied to mine datasets including various analytical signatures derived from the above-mentioned screening methods (Zielinski et al., 2014). Some examples of unsupervised methods are cluster analysis (CA) and principal component analysis (PCA), which have been utilised for classification and pre-process based on the FAs profiles of peanuts (Shin et al., 2010; Wang et al., 2017). Furthermore, several supervised algorithms and affiliated classifiers have been implemented, including linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), hierarchical cluster analysis (HCA), and partial least squares discriminant analysis (PLS-DA) (Ghosh et al., 2016; Granato, Santos, et al., 2018; Wang et al., 2014). Some regression methods for quantitative analysis have been applied as well, such as partial least squares (PLS), principal component regression (PCR), multiple linear regression (MLR), and support vector machine (SVM) (Govindarajan et al., 2009; Sun et al., 2019; Wei et al., 2015). In addition, some machine learning algorithms such as random forest (RF) are also applied for the model establishment for food classification (Menze et al., 2009). Finally, model validation is an essential stage since it enables an objective evaluation of whether the models are suitable for their targets (Alewijn et al., 2016).

1.5. KNOWLEDGE GAP

Peanuts as raw materials are the foundation of the whole peanut industry. From variety breeding, fertilisation, planting, harvesting, trading, and processing to consuming, each section in the peanut industry chain needs to be monitored to ensure the quality of peanuts. Although extensive research has been conducted on the trait evaluation of peanuts, the underlying distinct analytical signatures for the different peanut varieties, especially high oleic acid peanut varieties and single kernels for breeding, are not fully known. In addition, the demands for on-site rapid detection methods are rising. However, most research takes advantage of lab-based screening methods. The application of on-site screening methods to analyse the traits of peanuts has received limited attention so far.

On the other hand, the characteristics of products manufactured from different peanut varieties are imbalanced and scientific classification of peanut varieties based on processing utilisation can guarantee appropriate product characteristics for the envisaged manufacturing applications. Some research has already classified peanut varieties for manufacturing different types of peanut oil (Wang, 2016) and peanut protein (Wang et al., 2017). But in contrast, the impacts of raw materials on the characteristics of peanut butters are not completely known. Therefore, in order to ensure the reliability of peanut classification methods for processing stable peanut butter, it is necessary to elucidate the consistent distinctions between peanut butters from different varieties and underlying causes. Meanwhile, it would be useful to evaluate screening methods to identify these pivotal processing characteristics of peanuts.

1.6. RESEARCH AIM AND THESIS OUTLINE

The main aim of this thesis is to elucidate and comprehend distinct analytical signatures and relationships of various types of peanuts and derived peanut butters for quick evaluation and identification of peanuts to enhance the value of peanuts and final products in the whole production chain. The detailed objectives are:

(a) To understand the distinctions and similarities in the analytical signatures between different peanut varieties; to identify the quality traits of batch and single kernel measurements and to develop rapid screening approaches to assess the characteristics of peanuts (**Chapters 2** and **3**).

(b) To elucidate the relationships between peanuts and derived peanut butters; to classify peanuts systematically, and based on this, to develop rapid peanut identification methods for peanut butter manufacture (**Chapters 4** and **5**).

The framework of the thesis is shown in **Figure 1.3**. **Chapter 1** presents a general introduction of this thesis, including the background information and the scientific issues.

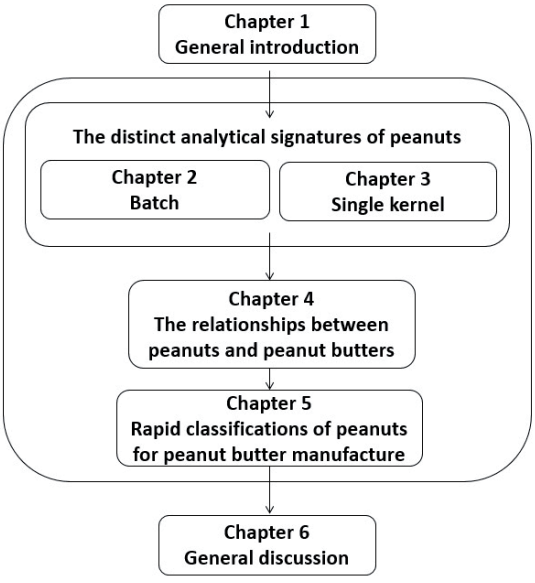


Figure 1.3. Framework of the thesis.

Chapter 2 explores the FAs signatures between high oleic acid peanuts and regular peanuts for the establishment of rapid batch measurements. The spectral features of different peanut types (n = 150) are examined by portable and benchtop NIRS. The differences in FAs compositions between different peanut types are also determined by gas chromatography, as well as the underlying causes for the spectral differences observed. The performance of classification models for distinguishing peanut types of different NIRS

analyser types is evaluated. Furthermore, the portable NIRS is assessed for its quantitative measurement capability of major FAs in comparison with a benchtop NIRS device.

Chapter 3 explores the analytical signatures of single peanut kernels (n = 110). The internal and external characteristics of different single peanuts are examined to optimise the design of a single detection accessory combined with portable NIRS for obtaining spectral data. The difference in compositional data between different peanut varieties is explored as well by principal component analysis. A robust partial least squares regression method is applied to build models for the prediction of macro components of individual peanuts. The performance of this method is estimated, and the spectral data associated with the structure of molecular groups of peanuts are pinpointed as well.

Chapter 4 focuses on the impacts of peanuts on the quality traits of peanut butters and explores their multivariate relationships. Peanut butters are processed from forty peanut varieties with comprehensive compositional analysis (fat, FAs, protein, amino acids, and sucrose of peanuts) and comparison. The volatile compounds, colour, texture, rheology, and particle size distributions of the corresponding peanut butters are comprehensively analysed. The heterogeneities between different varieties are evaluated with both univariate and multivariate statistical methods. The correlations between peanuts and peanut butters are investigated, and the underlying causes for the correlations are discussed.

Chapter 5 aims to develop a novel and rapid method to classify peanut varieties for the preparation of different types of peanut butter. Principal component analysis combined with cluster analysis of the structure characteristics (texture and rheology) and roast characteristics (colour and volatile compounds) of the resulting peanut butters is conducted to group peanut varieties, respectively. The latent causes for the spectral differences of the peanuts are analysed. Different spectral pre-treatments, modelling methods, and feature extraction methods are compared to develop robust classification models. The sensitivity, specificity, and accuracy of models and kernel density estimation are assessed to select the optimal methods.

Finally, the general discussion (**Chapter 6**) integrates the results from **Chapter 2** to **Chapter 5.** This chapter compares the analytical signatures of batch and single kernels and evaluates the screening methods. Furthermore, the impacts of raw materials on the quality traits of peanut butters and the peanut clustering results based on processed peanut butter application as well as rapid identification methods are assessed. Meanwhile, the comparison between the lab-based confirmatory and screening methods and the implications for the peanut chain are proposed. Finally, the limitations of this thesis and recommendations for further research are presented in this chapter.

REFERENCES

Agelet, L. E., & Hurburgh, C. R., Jr. (2014). Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. *Talanta, 121*, 288-299. https://doi.org/10.1016/j.talanta.2013.12.038.

Alewijn, M., van der Voet, H., & van Ruth, S. (2016). Validation of multivariate classification methods using analytical fingerprints – concept and case study on organic feed for laying hens. *Journal of Food Composition and Analysis, 51*, 15-23. https://doi.org/10.1016/j.jfca.2016.06.003.

And, C. M. L., & Resurreccion, A. V. A. (2002). Improved Correlation Between Sensory and Instrumental Measurement of Peanut Butter Texture. *Journal of Food Science, 67*(5), 1939-1949. https://doi.org/https://doi.org/10.1111/j.1365-2621.2002.tb08750.x.

Archer, P. (2016). Overview of the peanut industry supply chain. In *Peanuts* (pp. 253-266): Elsevier.

Bera, S. K., Kamdar, J. H., Kasundra, S. V., Patel, S. V., Jasani, M. D., Maurya, A., Dash, P., Chandrashekar, A. B., Rani, K., & Manivannan, N. (2019). Steady expression of high oleic acid in peanut bred by marker-assisted backcrossing for fatty acid desaturase mutant alleles and its effect on seed germination along with other seedling traits. *PLoS ONE, 14*(12), e0226252.

Bertioli, D. J., Seijo, G., Freitas, F. O., Valls, J. F., Leal-Bertioli, S. C., & Moretzsohn, M. C. (2011). An overview of peanut and its wild relatives. *Plant Genetic Resources, 9*(1), 134-149.

Bilal, M., Xiaobo, Z., Arslan, M., Tahir, H. E., Azam, M., Junjun, Z., Basheer, S., & Abdullah. (2020). Rapid determination of the chemical compositions of peanut seed (Arachis hypogaea.) Using portable near-infrared spectroscopy. *Vibrational Spectroscopy, 110*, 103138. https://doi.org/10.1016/j.vibspec.2020.103138.

Castro, R. C., Ribeiro, D. S. M., Santos, J. L. M., & Páscoa, R. N. M. J. (2021). Comparison of near infrared spectroscopy and Raman spectroscopy for the identification and quantification through MCR-ALS and PLS of peanut oil adulterants. *Talanta, 230*. https://doi.org/10.1016/j.talanta.2021.122373.

CBI. (2020). Entering the European market for groundnuts. Retrieved from: https://www.cbi.eu/market-information/processed-fruit-vegetables-edible-nuts/groundnuts/market-entry Accessed.

Cheng, J.-H., Jin, H., & Liu, Z. (2018). Developing a NIR multispectral imaging for prediction and visualization of peanut protein content using variable selection algorithms. *Infrared Physics & Technology, 88*, 92-96. https://doi.org/10.1016/j.infrared.2017.11.018.

Chetschik, I., Granvogl, M., & Schieberle, P. (2010). Quantitation of key peanut aroma compounds in raw peanuts and pan-roasted peanut meal. Aroma reconstitution and comparison with commercial peanut products. *Journal of Agricultural and Food Chemistry, 58*(20), 11018-11026. https://doi.org/10.1021/jf1026636.

Chukwumah, Y., Walker, L., Vogler, B., & Verghese, M. (2012). Profiling of bioactive compounds in cultivars of Runner and Valencia peanut market-types using liquid chromatography/APCI mass spectrometry. *Food Chemistry, 132*(1), 525-531.

Citerne, C. P., Pierna, J. C., & Moan, M. (2001). Rheological properties of peanut butter. *Rheologica Acta, 40*, 86-96. https://doi.org/10.1007/s003970000120.

Culbreath, A., Todd, J., Gorbet, D., Brown, S., Baldwin, J., Pappu, H., Holbrook, C., & Shokes, F. (1999). Response of early, medium, and late maturing peanut breeding lines to field epidemics of tomato spotted wilt. *Peanut Science, 26*(2), 100-106.

Davis, J. P., & Dean, L. L. (2016). Peanut composition, flavor and nutrition. In *Peanuts* (pp. 289-345): Elsevier.

de Oliveira Aparecido, L. E., Lorençone, J. A., Lorençone, P. A., de Meneses, K. C., & da Silva Cabral de Moraes, J. R. (2021). Climate risk to peanut cultivation in Brazil across different planting seasons. *Journal of the Science of Food and Agriculture*.

Deiana, P., Santona, M., Dettori, S., Culeddu, N., Dore, A., & Molinu, M. G. (2019). Multivariate approach to assess the chemical composition of Italian virgin olive oils as a function of variety and harvest period. *Food Chemistry, 300*, 125243. https://doi.org/10.1016/j.foodchem.2019.125243.

Dhamsaniya, N. K., Patel, N. C., & Dabhi, M. N. (2012). Selection of groundnut variety for making a good quality peanut butter. *Journal of Food Science and Technology, 49*(1), 115-118. https://doi.org/10.1007/s13197-011-0361-9.

Fan, D., Yan, B., Lian, H., Zhao, J., & Zhang, H. (2016). Instrumental and Sensory Analysis of the Properties of Traditional Chinese Fried Fritters. *Journal of Chemistry, 2016*, 1-7. https://doi.org/10.1155/2016/3626581.

Faostat, F. (2020). Statistical databases. Retrieved from: http://www.fao.org/faostat/en/#data/QC Accessed.

Farber, C., Sanchez, L., Rizevsky, S., Ermolenkov, A., McCutchen, B., Cason, J., Simpson, C., Burow, M., & Kurouski, D. (2020). Raman Spectroscopy Enables Non-Invasive Identification of Peanut Genotypes and Value-Added Traits. *Sci Rep, 10*(1), 7730. https://doi.org/10.1038/s41598-020-64730-w.

Fox, G., & Cruickshank, A. (2005). Near infrared reflectance as a rapid and inexpensive surrogate measure for fatty acid composition and oil content of peanuts (Arachis hypogaea L.). *Journal of Near Infrared Spectroscopy, 13*(5), 287.

GB5009.5. (2016). *National Food Safety Standard - Determination of Protein in Foods*. Beijing: Standard Press of China.

GB5009.6. (2016). *National Food Safety Standard - Determination of Fat in Foods*. Beijing: Standard Press of China.

GB5009.8. (2016). *National food safety standard - Determination of fructose, glucose, sucrose, maltose and lactose in foods* Beijing: Standard Press of China.

GB5009.124. (2016). *National Food Safety Standard - Determination of Amino Acid in Foods*. Beijing: Standard Press of China.

GB5009.168. (2016). National food safety standard- Determination of Fatty Acids in Foods *Beijing: Standard Press of China*.

Ghosh, S., Mishra, P., Mohamad, S. N. H., de Santos, R. M., Iglesias, B. D., & Elorza, P. B. (2016). Discrimination of peanuts from bulk cereals and nuts by near infrared reflectance spectroscopy. *Biosystems Engineering, 151*, 178-186.

Gong, A., Shi, A., Liu, H., Yu, H., Liu, L., Lin, W., & Wang, Q. (2018). Relationship of chemical properties of different peanut varieties to peanut butter storage stability. *Journal of Integrative Agriculture, 17*(5), 1003-1010. https://doi.org/10.1016/s2095-3119(18)61919-7.

Gorrepati, K., Balasubramanian, S., & Chandra, P. (2015). Plant based butters. *Journal of Food Science and Technology, 52*(7), 3965-3976. https://doi.org/10.1007/s13197-014-1572-7.

Govindarajan, K., Kandala, C., & Subbiah, J. (2009). NIR reflectance spectroscopy for nondestructive moisture content determination in peanut kernels. *Transactions of the ASABE, 52*(5), 1661-1666.

Granato, D., Putnik, P., Kovačević, D. B., Santos, J. S., Calado, V., Rocha, R. S., Cruz, A. G. D., Jarvis, B., Rodionova, O. Y., & Pomerantsev, A. (2018). Trends in Chemometrics: Food Authentication, Microbiology, and Effects of Processing. *Comprehensive reviews in food science and food safety, 17*(3), 663-677. https://doi.org/10.1111/1541-4337.12341.

Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L., & Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: A critical perspective. *Trends in Food Science & Technology, 72*, 83-90. https://doi.org/10.1016/j.tifs.2017.12.006.

Ha, J. W., Kim, S. Y., Ryu, S. R., & Kang, D. H. (2013). Inactivation of Salmonella enterica serovar Typhimurium and Escherichia coli O157:H7 in peanut butter cracker sandwiches by radio-frequency heating. *Food Microbiol, 34*(1), 145-150. https://doi.org/10.1016/j.fm.2012.11.018.

Hammons, R. O., Herman, D., & Stalker, H. T. (2016). Origin and early history of the peanut. In *Peanuts* (pp. 1-26): Elsevier.

Hashemian, M., Murphy, G., Etemadi, A., Dawsey, S. M., Liao, L. M., & Abnet, C. C. (2017). Nut and peanut butter consumption and the risk of esophageal and gastric cancer subtypes. *The American Journal of Clinical Nutrition, 106*(3), 858-864. https://doi.org/10.3945/ajcn.117.159467.

Haytowitz, D. B. A., Jaspreet K.C.; Wu, Xianli; Somanchi, Meena; Nickle, Melissa; Nguyen, Quyen A.; Roseland, Janet M.; Williams, Juhi R.; Patterson, Kristine Y.; Li, Ying; Pehrsson, Pamela R.. . (2021). *USDA national nutrient database for standard reference, Legacy Release*. https://data.nal.usda.gov/dataset/usda-national-nutrient-database-standard-reference-legacy-release.

Heras-Roger, J., Diaz-Romero, C., & Darias-Martin, J. (2016). A comprehensive study of red wine properties according to variety. *Food Chemistry, 196*, 1224-1231. https://doi.org/10.1016/j.foodchem.2015.10.085.

Huang, F., Li, Y., Guo, H., Xu, J., Chen, Z., Zhang, J., & Wang, Y. (2016). Identification of waste cooking oil and vegetable oil via Raman spectroscopy. *Journal of Raman Spectroscopy, 47*(7), 860-864. https://doi.org/10.1002/jrs.4895.

IS9037. (1979). *Peanut butter*. India: Bureau of Indian Standards

Jiang, J., Qiao, X., & He, R. (2016). Use of Near-Infrared hyperspectral images to identify moldy peanuts. *Journal of Food Engineering, 169*, 284-290. https://doi.org/10.1016/j.jfoodeng.2015.09.013.

Jin, H., Li, H., Yin, Z., Zhu, Y., Lu, A., Zhao, D., & Li, C. (2021). Application of Raman spectroscopy in the rapid detection of waste cooking oil. *Food Chemistry, 362*, 130191. https://doi.org/10.1016/j.foodchem.2021.130191.

Kilcast, D. (2013). 1 - Measurement of the sensory quality of food: an introduction. In D. Kilcast (Ed.), *Instrumental Assessment of Food Sensory Quality* (pp. 1-26): Woodhead Publishing.

Kundu, M., Khatkar, B. S., & Gulia, N. (2017). Assessment of chapatti quality of wheat varieties based on physicochemical, rheological and sensory traits. *Food Chemistry, 226*, 95-101. https://doi.org/10.1016/j.foodchem.2016.12.046.

Lan, W., Jaillais, B., Leca, A., Renard, C., & Bureau, S. (2020). A new application of NIR spectroscopy to describe and predict purees quality from the non-destructive apple measurements. *Food Chemistry, 310*, 125944. https://doi.org/10.1016/j.foodchem.2019.125944.

Liedl Jr, F. G., & Rowe, K. F. (2007). Nut butter and related products and method of making same. Google Patents.

Lou, F., Liu, Y., Sun, X., Pan, Y., Zhao, J., & Zhao, Y. (2009). Identification of Volatile Flavor Components in Peanut Butter *Food Science, 24*, 393-396.

Ma, L., Li, B., Han, F., Yan, S., Wang, L., & Sun, J. (2015). Evaluation of the chemical quality traits of soybean seeds, as related to sensory attributes of soymilk. *Food Chemistry, 173*, 694-701. https://doi.org/10.1016/j.foodchem.2014.10.096.

Menze, B. H., Kelm, B. M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., & Hamprecht, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics, 10*(1), 1-16.

Mohd Rozalli, N. H., Chin, N. L., & Yusof, Y. A. (2015). Particle Size Distribution of Natural Peanut Butter and Its Dynamic Rheological Properties. *International Journal of Food Properties, 18*(9), 1888-1894. https://doi.org/10.1080/10942912.2014.971184.

Mohd Rozalli, N. H., Chin, N. L., Yusof, Y. A., & Mahyudin, N. (2016). Quality changes of stabilizer-free natural peanut butter during storage. *Journal of Food Science and Technology, 53*(1), 694-702. https://doi.org/10.1007/s13197-015-2006-x.

Norazatul Hanim, M. R., Chin, N. L., & Yusof, Y. A. (2016). Effects of Grinding Time on Rheological, Textural and Physical Properties of Natural Peanut Butter Stored at Different Temperatures. *Journal of Texture Studies, 47*(2), 131-141. https://doi.org/10.1111/jtxs.12167.

NY/T958. (2006). *Peanut butter*. Beijing: The Ministry of Agriculture and Rural Affairs.

Pérez-Magariño, S., Ortega-Heras, M., Bueno-Herrera, M., Martínez-Lapuente, L., Guadalupe, Z., & Ayestarán, B. (2015). Grape variety, aging on lees and aging in bottle after disgorging influence on volatile composition and foamability of sparkling wines. *LWT - Food Science and Technology, 61*(1), 47-55. https://doi.org/10.1016/j.lwt.2014.11.011.

Phan-Thien, K.-Y., Golic, M., Wright, G. C., & Lee, N. A. (2011). Feasibility of estimating peanut essential minerals by near infrared reflectance spectroscopy. *Sensing and Instrumentation for Food Quality and Safety, 5*(1), 43-49. https://doi.org/10.1007/s11694-011-9110-z.

Qiao, X., Jiang, J., Qi, X., Guo, H., & Yuan, D. (2017). Utilization of spectral-spatial characteristics in short-wave infrared hyperspectral images to classify and identify fungi-contaminated peanuts. *Food Chemistry, 220*, 393-399. https://doi.org/10.1016/j.foodchem.2016.09.119.

Rao, Y., Xiang, B., Zhou, X., Wang, Z., Xie, S., & Xu, J. (2009). Quantitative and qualitative determination of acid value of peanut oil using near-infrared spectrometry. *Journal of Food Engineering, 93*(2), 249-252.

Shakerardekani, A., Karim, R., Ghazali, H. M., & Chin, N. L. (2013). Textural, rheological and sensory properties and oxidative stability of nut spreads–a review. *International Journal of Molecular Sciences, 14*(2), 4223-4241. https://doi.org/10.3390/ijms14024223.

Shin, E.-C., Craft, B. D., Pegg, R. B., Phillips, R. D., & Eitenmiller, R. R. (2010). Chemometric approach to fatty acid profiles in Runner-type peanut cultivars by principal component analysis (PCA). *Food Chemistry, 119*(3), 1262-1270. https://doi.org/10.1016/j.foodchem.2009.07.058.

SRD. (2020). Peanut Butter Industry - Statistics & Facts. Retrieved from: https://www.statista.com/topics/2287/peanut-butter-industry/ Accessed.

Sun, J., Shi, X., Zhang, H., Xia, L., Guo, Y., & Sun, X. (2019). Detection of moisture content in peanut kernels using hyperspectral imaging technology coupled with chemometrics. *Journal of Food Process Engineering, 42*(7), e13263.

Sun, J., Wang, G., Zhang, H., Xia, L., Zhao, W., Guo, Y., & Sun, X. (2020). Detection of fat content in peanut kernels based on chemometrics and hyperspectral imaging technology. *Infrared Physics & Technology, 105*. https://doi.org/10.1016/j.infrared.2020.103226.

Talcott, S. T., Passeretti, S., Duncan, C. E., & Gorbet, D. W. (2005). Polyphenolic content and sensory properties of normal and high oleic acid peanuts. *Food Chemistry, 90*(3), 379-388. https://doi.org/10.1016/j.foodchem.2004.04.011.

Tanti, R., Barbut, S., & Marangoni, A. G. (2016). Oil stabilization of natural peanut butter using food grade polymers. *Food Hydrocolloids, 61*, 399-408. https://doi.org/10.1016/j.foodhyd.2016.05.034.

Tillman, B. L., Gorbet, D. W., & Person, G. (2006). Predicting oleic and linoleic acid content of single peanut seeds using near-infrared reflectance spectroscopy. *Crop Science, 46*(5), 2121-2126.

Toomer, O. T. (2018). Nutritional chemistry of the peanut (Arachis hypogaea). *Critical Reviews in Food Science and Nutrition, 58*(17), 3042-3053.

USDA. (1972). United States Standards for Grades of Peanut Butter. Retrieved from: https://www.ams.usda.gov/grades-standards/peanut-butter-grades-and-standards Accessed.

Venkatachalam, M., & Sathe, S. K. (2006). Chemical composition of selected edible nut seeds. *Journal of Agricultural and Food Chemistry, 54*(13), 4705-4714.

Wang, L., Liu, H., Liu, L., Wang, Q., Li, Q., Du, Y., & Zhang, J. (2014). Protein contents in different peanut varieties and their relationship to gel property. *International Journal of Food Properties, 17*(7), 1560-1576.

Wang, L., Liu, H., Liu, L., Wang, Q., Li, S., & Li, Q. (2017). Prediction of peanut protein solubility based on the evaluation model established by supervised principal component regression. *Food Chemistry, 218*, 553-560. https://doi.org/10.1016/j.foodchem.2016.09.091.

Wang, L., Wang, Q., Liu, H., Liu, L., & Du, Y. (2013). Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy. *Journal of Food Science and Technology, 93*(1), 118-124. https://doi.org/10.1002/jsfa.5738.

Wang, Q. (2016). *Peanuts: Processing Technology and Product Development*: Academic Press.

Wang, Q. (2018). Overview of peanut processing quality. In *Peanut Processing Characteristics and Quality Evaluation* (pp. 1-67): Springer.

Wang, Q. (2020). *The database of peanut processing characteristics and special varieties*. National Science & Technology Infrastructure of China, agridata.cn/data.html#/datadetail?id=286518.

Wei, Z., Wang, J., & Zhang, W. (2015). Detecting internal quality of peanuts during storage using electronic nose responses combined with physicochemical methods. *Food Chemistry, 177*, 89-96.

Wu, D., Chen, X., Shi, P., Wang, S., Feng, F., & He, Y. (2009). Determination of alpha-linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination. *Analytica Chimica Acta, 634*(2), 166-171. https://doi.org/10.1016/j.aca.2008.12.024.

Yang, M., Chen, Q., Kutsanedzie, F. Y. H., Yang, X., Guo, Z., & Ouyang, Q. (2017). Portable spectroscopy system determination of acid value in peanut oil based on variables selection algorithms. *Measurement, 103*, 179-185. https://doi.org/10.1016/j.measurement.2017.02.037.

Yang, Y., Zhang, L., Hettinga, K. A., Erasmus, S. W., & van Ruth, S. M. (2020). Prevalence of Milk Fraud in the Chinese Market and its Relationship with Fraud Vulnerabilities in the Chain. *Foods (Basel, Switzerland), 9*(6), 709. https://doi.org/10.3390/foods9060709.

Yu, H., Liu, H., Wang, N., Yang, Y., Shi, A., Liu, L., Hu, H., Mzimbiri, R. I., & Wang, Q. (2016). Rapid and visual measurement of fat content in peanuts by using the hyperspectral imaging technique with chemometrics. *Analytical Methods, 8*(41), 7482-7492. https://doi.org/10.1039/c6ay02029a.

Yu, H. W., Wang, Q., Shi, A. M., Yang, Y., Liu, L., Hu, H., & Liu, H. Z. (2017). Visualization of Protein in Peanut Using Hyperspectral Image with Chemometrics. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis, 37*(3), 853-858. https://doi.org/10.3964/j.issn.1000-0593(2017)03-0853-06.

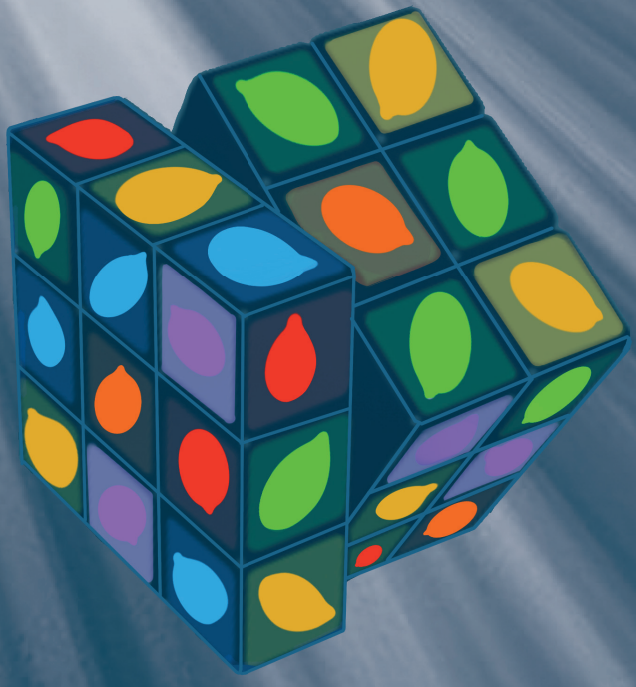
Yuan, D., Jiang, J., Qiao, X., Qi, X., & Wang, W. (2020). An application to analyzing and correcting for the effects of irregular topographies on NIR hyperspectral images to improve identification of moldy peanuts. *Journal of Food Engineering, 280*. https://doi.org/10.1016/j.jfoodeng.2020.109915.

Zeng, L., Tu, B., Yin, C., Zheng, X., Song, Z., He, D., & Qi, P. (2016). Qualitative and quantitative analysis of peanut oil adulteration by laser near infrared spectroscopy with SVM. *Journal of the Chinese Cereals and Oils Association, 31*(8), 126-130 and 137. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85017369447&partnerID=40&md5=d8a5a26da08cdf8ba236758a12db5eb7.

Zhang, J. (2012). The effects of fatty acid compositions and endogenous antioxidants of peanut cultivars on oil stability. Beijing, China: Chinese Academy of Agricultural Sciences.(in Chinese).

Zhao, Z., Shi, A., Wang, Q., & Zhou, J. (2019). High Oleic Acid Peanut Oil and Extra Virgin Olive Oil Supplementation Attenuate Metabolic Syndrome in Rats by Modulating the Gut Microbiota. *Nutrients, 11*(12). https://doi.org/10.3390/nu11123005.

Zielinski, A. A., Haminiuk, C. W., Nunes, C. A., Schnitzler, E., van Ruth, S. M., & Granato, D. (2014). Chemical composition, sensory properties, provenance, and bioactivity of fruit juices as assessed by chemometrics: a critical review and guideline. *Comprehensive reviews in food science and food safety, 13*(3), 300-316.



CHAPTER 2

Evaluation of portable and benchtop near-infrared spectroscopy for classification of high oleic acid peanuts and fatty acid quantitation

This chapter has been published as:
Yu, H., Liu, H., Wang, Q., & van Ruth, S. M. (2020). Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. LWT - Food Science and Technology, 128, 109398.

ABSTRACT

Portable near-infrared spectroscopy (NIRS) analyser for classifying the high oleic acid peanuts (HOP) and quantitation of their major fatty acids was assessed for the first time in comparison with the benchtop NIRS. Reference chemical values of fatty acids were calculated by the gas chromatographic method. The processed datasets were explored by principal component analysis and classification models were built by using partial least squares discriminant analysis. The results showed that the accuracy of distinction of the HOP from others was 100%. Partial least squares analysis was used to build quantitative models for quantifying the peanuts' major fatty acids. The R of the calibration model noted for the portable NIRS was 0.90, 0.88, and 0.88 for oleic acid, linoleic acid, and palmitic acid of the HOP with a SEC of 0.97, 0.12, and 0.12, respectively. The similar results could be found in the benchtop NIRS. The RPD of all models were over 2 which showed good performance of the models. This study indicated that the portable NIRS performance was comparable with the performance of the benchtop NIRS for distinction of the HOP from others, as well as for the prediction of the contents of their main fatty acids.

Keywords: High oleic acid; Partial least squares discriminant analysis; Partial least squares; Peanut; Portable near-infrared spectroscopy analyser

2.1. INTRODUCTION

High oleic acid peanuts (HOP) are increasingly used in product processing and gradually become the main raw material worldwide because they have several advantages over the regular peanuts (RP), such as an extended shelf time for its products (Gong et al., 2018; Olmedo et al., 2018) and an enhanced health value for human beings (Suchoszek-Łukaniuk et al., 2011; Vassiliou et al., 2009). Obviously, the main differences between HOP and RP are their fatty acid contents. HOP have higher oleic acid, lower linoleic acid, and lower palmitic acid contents (Nepote et al., 2006). However, RP are easily pretending to be high oleic peanuts because there is no difference in appearance. The farmers, processors, and consumers are vulnerable to be cheated with regular peanuts without any effective means of identification. The conventional method, gas chromatography (GC), for detecting fatty acid is time-consuming and out of reach for farmers and consumers. Therefore, some new methods should be selected to satisfy the efficient testing requirements instead of the chemical approach to ensure the rights of farmers and consumers.

Currently, rapid analysis methods based on physical traits like optical, electrical, and acoustical characteristics have been widely applied to the agriculture products (Xiaobo et al., 2016). Among them, near-infrared spectroscopy (NIRS), based on overtones and combined frequency absorption of vibration of hydrogen-containing groups (X-H), has been widely used in fatty acids analysis such as milk (Liu et al., 2018), cheese (Soto-Barajas et al., 2013) , salmon oil (Cascant et al., 2018), and maize (Egesel et al., 2016). The quality of peanuts and its product have been also examined by NIRS, such as for its fat and protein contents (Yu et al., 2016; Yu et al., 2017), and for fatty acid analysis in peanuts (Tillman et al., 2006) and peanut oil (Wu et al., 2009). However, these methods cannot give a good result when RP and HOP appear at the same time because there is no model that covers the whole range of oleic acid and linoleic acid. In order to get a better result, qualitative analysis for selecting HOP should precede quantitative analysis. Meanwhile, most of the above research adopted the benchtop device, which need relatively stable surroundings, usually a laboratory environment. With optical technology developing, the optical elements have been miniaturised with reliable performance to observably improve the portability of NIRS devices. Previously, studies have been conducted to adopt the portable NIRS for compositional analysis like acid value determination (Yang et al., 2017), and for classification such as fruit and tea quality assessment (Cirilli et al., 2016; Feng et al., 2019; Malegori et al., 2017; Wang et al., 2019), extra virgin olive oil distinction (Yan et al., 2019), and UHT milk classification (de Lima et al., 2018). Good results were obtained by using portable NIRS combined with advanced chemometrics in these studies. The purpose of this study is to distinguish HOP from RP and to quantify their major fatty acids by portable NIRS, and compare results with those of benchtop NIRS.

2.2. MATERIALS AND METHODS

2.2.1. Peanut Samples

150 different peanut varieties and strains from 10 main planting provinces like Shandong, Hebei, Henan province in China were offered by breeders from the National Peanut Industry System. The ratio of oleic acid to linoleic acid (O/L value) is more than 9:1 which is the minimum requirement of HOP according to Nawade study (Nawade et al., 2018). By contrast, for RP, the O/L value is normally less than 2:1. The sample set included 75 HOP samples and 75 RP samples, which both covered the main planting areas. These in-shelled peanuts were kept at 4 °C in refrigerated storage (Yuandong Co., LTD., Tianjin, China) after collection. Each sample with skin was shelled and analysed by a portable and benchtop NIRS immediately, and then fatty acid analysis was carried out consecutively. The workflow of this study was presented in **Figure 2.1**.

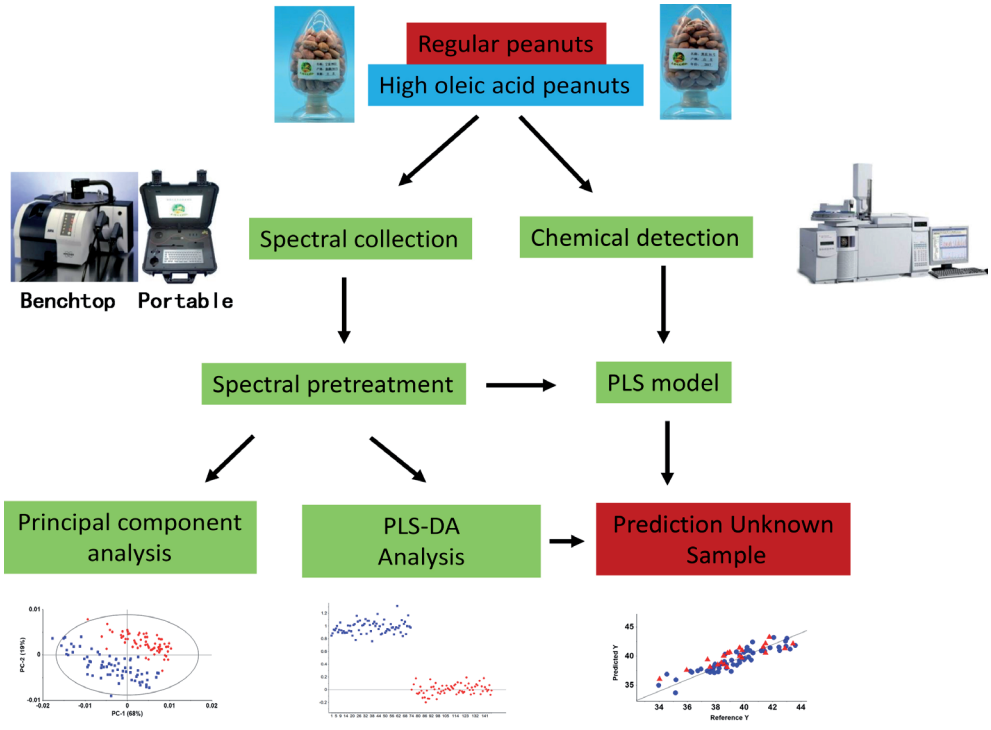


Figure 2.1. Workflow of the study. (PLS) Partial least squares; (PLS-DA) Partial least squares discriminant analysis.

2.2.2. Benchtop NIRS

A multipurpose-analyser (MPA) benchtop spectrometer (Bruker Corporation, Germany) was applied to collect the spectral data. 100 g peanut kernels were added to the sample cup (height 45 mm; diameter 97 mm) for each measurement based on a rotating sphere. The light through Quartz glass at the bottom of sample cup detected the peanut kernels, and the reflectance light was collected by a sensitive indium gallium arsenide (InGaAs) detector located inside the device. The spectra were acquired using OPUS 7.5 software (Bruker Corporation, Germany), over the range from 950 to 1650 nm with a resolution of 8 cm^{-1}. Each spectrum was the average of 32 scans and each sample was analysed in triplicate. The average of triplicate was calculated for chemometric analysis.

2.2.3. Portable NIRS

A box-portable NIRS was designed for easy transport and used by our research team. The system is based on a Micro-NIRS 1700 device (VIAVI Solutions, the USA) with a spectral range of 908 to 1676 nm and a 6 nm sampling interval. The device collected the spectral data in transmission mode and with the linear variable filter (LVF) applied. The box-portable NIRS analyser is controlled by a tablet (Surface, Microsoft Corporation, the USA), and the reference white board is Teflon. This equipment is powered by a lithium battery which would power the system for 12 h. The sample cup (height 50 mm; diameter 51 mm) was filled with about 60 g of peanut kernels for each measurement which would consist of 32 scans. This procedure was repeated five times. Measurements were carried out after dark and white calibration. Sample data were averaged for chemometric analysis. The spectra were collected by using Micro-NIRS Pro 2.4 software (VIAVI Solutions, the USA).

2.2.4. Fatty acids analysis

The fatty acids of the peanut samples were determined on a GC-2010 (Shimadzu, Japan) according to AOAC standard 996.01(Robinson et al., 2008). Fatty acid methyl esters (FAMEs) were determined by GC equipped with a 100 m × 0.25 mm ×0.20 μm Supelco SP-2560 column (Shimadzu, Japan). The detector was a flame ionisation and Helium the carrier gas. All reagents were ACS grade and acquired from Fisher Scientific (Thermo Scientific, the USA). Based on AOAC (1990) method, the fat contents in peanut (5 g) were extracted with mineral ether by Soxhlet apparatus. After extraction, the reagent was volatilized by a rotary vacuum evaporator at 50 °C. A volume of 0.2 g peanut oil was weighted in a 30 mL sterile, screw top plastic bottle. To start the transesterification, 2 mL sulfuric acid-methanol solution was added, heating it in a water bath at 70 °C for 1 h and shaking once every 20 min. After that, 2 mL of hexane was added to the mixture and then add distilled water until the water was flush with the tube bottleneck. The upper hexane

phase was passed through anhydrous sodium sulfate to remove moisture, and the resulting solution was used for GS analysis. The FAMEs were analysed based on retention times using standard substance, and the content of each fatty acid was calculated according to the peak area. Each sample was measured in duplicate and the average values were calculated for chemometric analysis.

2.2.5. Chemometric analysis

The oleic acid, linoleic acid, and palmitic acid concentrations of HOP and RP were employed as the dependent variables (*Y*) for chemometric analysis, whilst the collected spectra of portable NIRS and benchtop NIRS were the independent variables (*X*), respectively. The following statistics and chemometric analysis were applied in sequence, as follows: 1) boxplot analysis (Quintelas et al., 2019), Levene's test and Kolmogorov-Smirnov test (Dettori et al., 2018) to determine *Y* outliers, homogeneity and ascertain normal distributions; 2) significance analysis by *t*-test; 3) principal component analysis (PCA) (Liu et al., 2019) to analyse different peanut types interrelationships and *X* outliers; 4) linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and partial least squares-discriminant analysis (PLS-DA) to classify the HOP and RP; 5) partial least squares (PLS) to develop the models for each fatty acid of HOP and RP by different equipment.

A boxplot analysis as the first test was carried out, and a box graph was given with the median as the middle, the 25th and 75th percentiles as the edges, and the maximum and minimum value as the whiskers (Quintelas et al., 2019). The whiskers were identified as a result of the interquartile distance between the 25th and 75th percentiles. Obviously, outliers were plotted individually and could be identified by visual inspection. Next, the Kolmogorov-Smirnov test was applied to ascertain the null hypothesis that *X* dataset is normally distributed, and Levene's test was used for homogeneity. A *t*-test was used for significance test and fatty acids with $P < 0.05$ were marked as significant different.

In order to increase the signal-to-noise ratio and to explore more useful information, all spectral datasets were pretreated in different ways, including the first derivative (FD), the second derivative (SD), standard normal variate (SNV), detrending (DT), baseline (BL), and combined pretreatment such as SD+DT, FD+DT, etc. After pretreatment, the spectral data were used for PCA, qualitative and quantitative analysis. The relationship between NIRS spectra and fatty acids was decided by calculating the correlation coefficient between fatty acids concentrations and wavelength absorbances.

Classification of peanuts in terms of oleic acid concentrations was determined by creating three different discriminant models: LDA, QDA, and PLS-DA. These models are supervised learning according to the relationship between spectral information and pea-

nut categories (Cascant et al., 2018). To be specific, spectral data were regarded as *X* variables in line with the above and the two known categories (HOP and RP) were *Y* variables. For PLS-DA, the *Y* variables were an inconsecutive numerical value (zero for HOP and one for RP), whilst LDA and QDA estimated classification values based on allocating different peanut types. The performance of discrimination models was evaluated based on several indicators, including: accuracy, the overall rate of correct classification; sensitivity, the rate of correct identification; and specificity, the rate of correct rejection (Liu et al., 2018).

PLS (Wu & Sun, 2013) as the classic linear regression modelling algorithm, was used to gain robust model. After that, validation was an indispensable step to evaluate the performance of model. External verification (one-third random sample of the sample sets) was respectively adopted to assess the robustness of the calibration model. Assessment parameters were the correlation coefficient (R) of regression in calibration and prediction, and standard error in calibration (SEC) and prediction (SEP) (Agyekum et al., 2019; Lin et al., 2014; Sheng et al., 2019). The best models had high R (1 is optimal) and low SEC/P. Other statistic such as the residual predictive deviation (RPD), calculated as the ratio between the standard deviation (SD) of the prediction set to the SEP, where RPD lower than 1.5 was regarded as improper model while the other ones with RPD greater than 2 were identified as excellent (Ma et al., 2019; Ncama et al., 2017; Ye et al., 2016). To avoid the model overfitting, the optimal PLS factor was selected according to the prediction residual error sum of square (PRESS). All the above analyses were performed in R 3.5.3 (R Foundation for Statistical Computing, Austria) and Unscrambler 10.3 (CAMO Software AS, Norway).

2.3. RESULTS AND DISCUSSION

2.3.1. Chemical data of RP and HOP

The results of fatty acid analysis of the peanut samples showed that oleic acid, linoleic acid, and palmitic acid were the major fat components in peanuts, accounting for 92% of total fatty acids, which was consistent with the previous research (Qiang, 2018). All peanut samples were analysed for the major fatty acids as *Y* datasets. The box-plot showed the distributions of the fatty acids (**Figure 2.2**). For the *Y* datasets, only three outliers were marked. The range of oleic acid, linoleic acid, palmitic acid for high oleic acid peanuts is 33.97-43.61 g/100 g, 1.15-2.48 g/100 g, and 2.18-3.56 g/100 g, respectively. Sequentially, a Kolmogovov-Smirnov analysis was applied to judge whether the *Y* datasets (oleic acid, linoleic acid, and palmitic acid) were normally distributed, and Levene's test was used for homogeneity. All data satisfied the normal distribution and equality of variances requirements. According to the results of *t*-test (**Table 2.1**), three fatty acids presented at significantly different levels in HOP and RP.

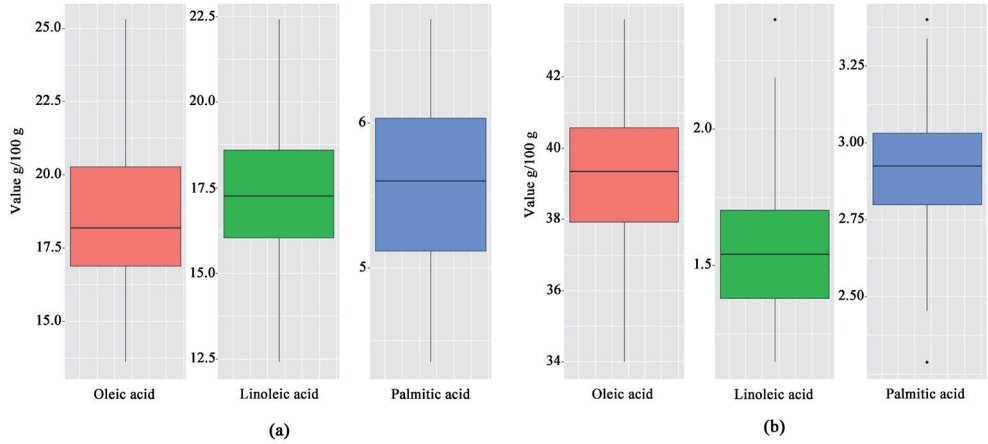


Figure 2.2. Distributions of the major fatty acids (a regular peanuts; b high oleic acid peanuts).

Table 2.1. The statistical analysis (mean concentrations, standard deviations, and *P* values) of the major fatty acids in regular/ high oleic acid peanuts (g/100 g peanut kernel).

Fatty acids		RP	HOP	P^{x}
Palmitic acid	C16:0	5.58 ± 0.60^{a}	2.92 ± 0.25^{b}	$<0.001^{***}$
Oleic acid	C18:1 n9c	18.76 ± 2.69^{a}	39.31 ± 2.25^{b}	$<0.001^{***}$
Linoleic acid	C18:2 n6c	17.48 ± 2.21^{a}	1.58 ± 0.27^{b}	$<0.001^{***}$

^{x} superscripts and an asterisk (***) indicate significant differences (*t*-test, $P < 0.001$). (HOP) high oleic acid peanuts; (RP) regular peanuts.

2.3.2. PCA models and correlation of spectral data and fatty acids

The upper of **Figure 2.3** shows the raw spectral data of different types of peanut acquired by two devices. In order to have the roundup of the attributes of different types of peanut, the PCA model combined with the best spectral pretreatment methods, was used for the data of the portable and benchtop NIRS. The optimised preprocessing methods for raw spectrum of portable and benchtop is FD and SNV transformation respectively. The total explain variance of PC1 and PC2 is 81% and 78% for portable NIRS and benchtop NIRS, which covers most of the information. The score distribution map of all peanut samples is showed in the middle of **Figure 2.3**, which reveals that HOP (red one) and RP (blue one) are relatively well separated. The PCA analysis revealed some outliers based on the Hotelling's T^2 (Liu et al., 2018). The significance level is 0.5% and the black ellipse is the limit. If points are located outside the black ellipse, these are the outliers.

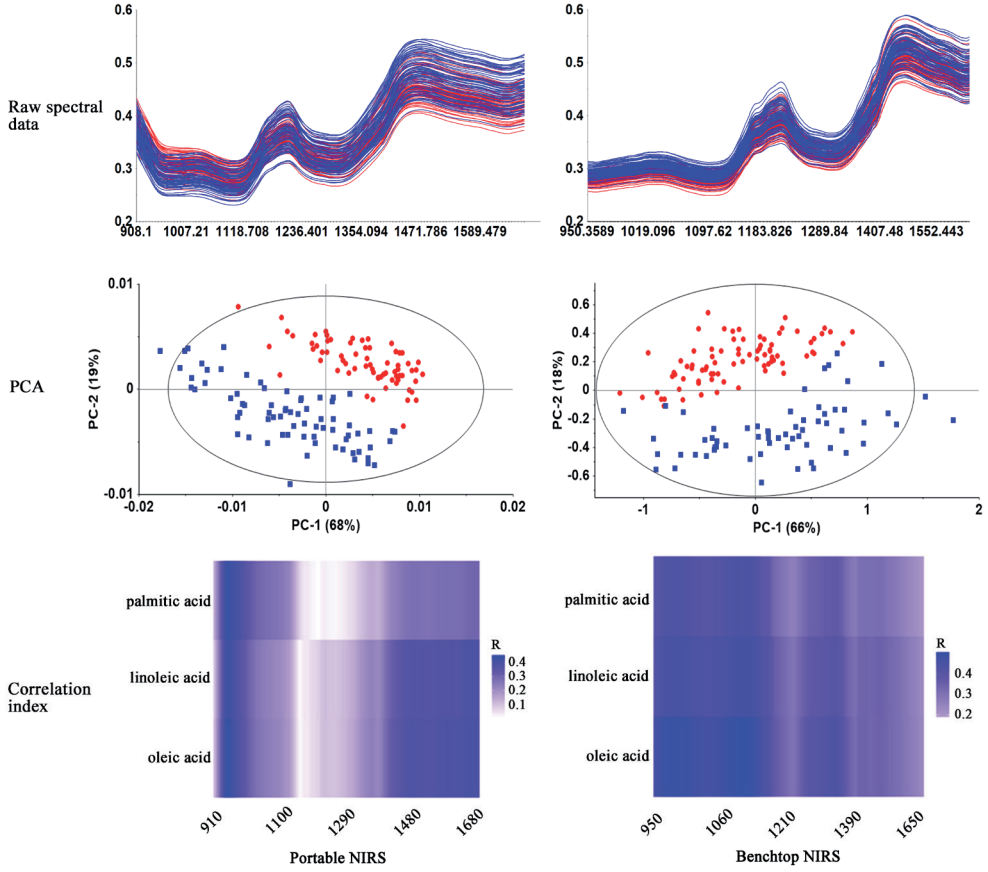


Figure 2.3. The raw spectral data (upper), the first two dimensions of principal component analysis (PCA, middle), and correlation plots of spectral data and the three main fatty acids (lower) for portable (left) and benchtop NIRS (right). High oleic acid peanuts are blue and regular peanuts are red. (NIRS) Near-infrared spectroscopy; (PCA) principal component analysis.

The lower of **Figure 2.3** shows the relationship between the spectral variations acquired by portable NIRS and benchtop NIRS (abscissa), and the fatty acids acquired by GC (ordinate). The portable NIRS spectra present predominantly correlation with the concentrations of oleic acid and linoleic acid in the spectral ranges from 930 to 960 nm and 1480 to 1680 nm, while it is from 930 to 1000 nm for the concentrations of palmitic acid. Palmitic acid is a highly abundant long chain saturated fatty acid, whereas oleic acid and linoleic acid are the unsaturated fatty acids. Similar results are found for benchtop NIRS. For the benchtop NIRS spectral data, FAs showing higher correlation coefficient values with wavelength ranges (950 to 1150 nm) and lower correlation values with wavelength ranges (1400 to 1650 nm) than portable NIRS due to equipment specific characteristics such as different optical path lengths (García Martín, 2015) and the principle of filter splitting.

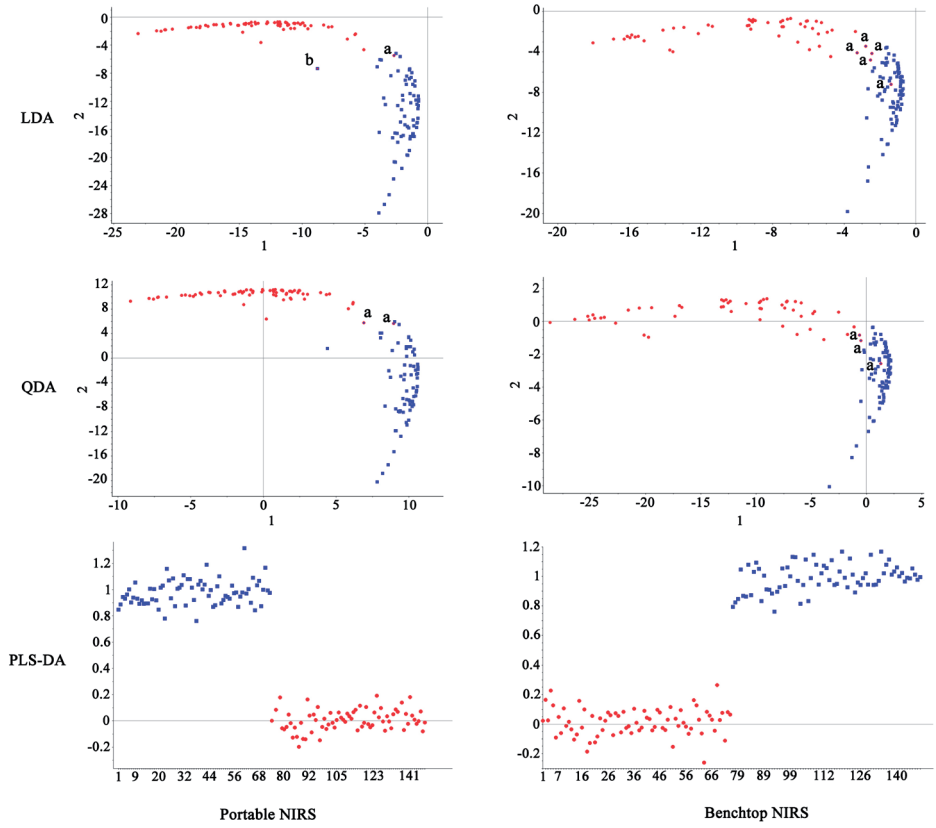


Figure 2.4. Discrimination plots of the LDA, QDA, and PLS-DA models to classify HOP (red) and RP (blue) (a, HOP be mistaken for RP; b, RP be mistaken for HOP). (HOP) High oleic acid peanuts; (LDA) Linear discriminant analysis; (NIRS) Near-infrared spectroscopy; (QDA) Quadratic discriminant analysis; (PLS-DA), Partial least squares-discriminant analysis; (RP) Regular peanuts.

2.3.3. Classification models

The possibility of classifying HOA and RP was assessed by developing and comparing three classification methods: LDA, QDA, and PLS-DA. As described above, a PCA model was built firstly in order to remove the irrelevant variables for LDA and QDA. The first two PCs explained 80 % of the spectral data. Therefore, these two components were adopted for constructing LDA and QDA models. The results about sensitivity, specificity, and accuracy of all the models obtained for the datasets are presented in **Table 2.2**. As shown in the table, only two or three samples were not accurately classified for the LDA and QDA for portable NIRS, while there were five and three wrong samples for benchtop NIRS. By contrast, all peanut samples were accurately classified by PLS-DA. The classification models of HOP (red one) and RP (blue one) are presented in **Figure 2.4**. In addition, the corresponding sensitivity, specificity, and accuracy values for each model are also shown in

Table 2.2. Sensitivity (SENS), specificity (SPEC), and accuracy (ACCU) of the discrimination models by portable NIRS and benchtop NIRS.

Index		n	Category		Portable NIRS (%)			Category		Benchtop NIRS (%)		
			HOP	RP	SENS	SPEC	ACCU	HOP	RP	SENS	SPEC	ACCU
LDA	HOP	73	72	1	98	98	98	68	5	93	100	97
	RP	73	1	72	98	98		0	73	100	94	
QDA	HOP	73	71	2	97	100	98	70	3	97	100	97
	RP	73	0	73	100	97		0	73	100	96	
PLS-DA	HOP	73	73	0	100	100	100	73	0	100	100	100
	RP	73	0	73	100	100		0	73	100	100	

(HOP) High oleic acid peanuts; (LDA) Linear discriminant analysis; (n) Number; (NIRS) Near-infrared spectroscopy; (QDA) Quadratic discriminant analysis; (PLS-DA) Partial least squares-discriminant analysis; (RP) Regular peanuts.

Table 2.2. The best performance derived from the PLS-DA model (sensitivity, specificity, and accuracy of 100%), although the performance of the LDA and QDA models were also acceptable (sensitivity, specificity, and accuracy higher than 93%). As a linear discrimination algorithm, PLS-DA has more robust than non-linear algorithms when analysing the multi-collinearity data (Gromski et al., 2015). In general, it can be confirmed that all the discriminant models achieved acceptable HOP classification according to spectral data whether it is a portable device or a benchtop one.

2.3.4. PLS regression

In order to predict the major fatty acids in HOP and RP, calibration models based on the chemical values and NIRS spectra were constructed by using PLS. The R, SEC/SEP, and RPD were chosen to evaluate the best model of each fatty acid. When constructing the calibration models, it was required to preprocess the raw NIRS spectra data because there were some interferences such as baseline drift and certain stray light effects. The preprocessing algorithms, as described above, are often used to reduce the interferences (Ye et al., 2016).

For different fatty acids of different NIRS equipment, the best processing algorithms were not the same. The best pretreatments for constructing the PLS models for fatty acids of different peanut types and the performances of each calibration and validation model are listed in **Table 2.3**. Overall speaking, the models of portable NIRS are similar to the models of benchtop NIRS for all the studied fatty acids. The best correlation coefficients were equal or above 0.88. The best pretreatment for oleic acid, linoleic acid, and palmitic acid of HOP based on portable NIRS was SD+DT, DT, and SD respectively. By contrast, for benchtop NIRS, the best pretreatment for oleic acid, linoleic acid, and palmitic acid of HOP was SNV, SD, and SD individually. Although the R of oleic acid of calibration and validation model of benchtop NIRS for HOP was higher than portable NIRS, the RMSE were almost

Table 2.3. Best model of each compound based on portable NIRS and benchtop NIRS.

Index	Fatty acids	Pretreatment	Factors	Calibration		Validation		
				R	SEC	R	SEP	RPD
Portable NIRS	Oleic acid (HOP)	SD+DT	9	0.90	0.97	0.87	1.10	2.12
	Linoleic acid (HOP)	DT	6	0.88	0.12	0.86	0.13	2.01
	Palmitic acid (HOP)	SD	5	0.88	0.12	0.87	0.10	2.40
	Oleic acid (RP)	FD+DT	11	0.90	1.27	0.87	1.31	2.05
	Linoleic acid (RP)	DT+SD	9	0.92	0.96	0.75	1.14	2.05
	Palmitic acid (RP)	SD+DT	12	0.92	0.26	0.88	0.34	2.29
Benchtop NIRS	Oleic acid (HOP)	SNV	11	0.94	0.77	0.89	1.06	2.20
	Linoleic acid (HOP)	SD	4	0.90	0.12	0.90	0.11	2.38
	Palmitic acid (HOP)	SD	6	0.94	0.10	0.88	0.11	2.18
	Oleic acid (RP)	FD+SNV	7	0.93	0.98	0.87	1.10	2.44
	Linoleic acid (RP)	DT+FD	8	0.91	0.96	0.86	1.04	2.25
	Palmitic acid (RP)	SD+DT	10	0.90	0.34	0.92	0.30	2.60

(DT) Detrending; (FD) The first derivative; (HOP) High oleic acid peanuts; (NIRS) Near-infrared spectroscopy; (PLS) Partial least squares; (R) Correlation coefficient; (RP) Regular peanuts; (RPD) Residual predictive deviation; (SD) The second derivative; (SEC) Standard error in calibration; (SEP) Standard error in prediction; (SNV) Standard normal variate.

equal, especially for SEP. The difference between the SEP is only 0.04. Similar results were obtained for linoleic acid and palmitic acid of HOP. The gap of SEP for linoleic acid and palmitic acid between portable and benchtop NIRS were 0.02 and 0.01 respectively. The RPD of all models were over 2, which means the prediction models have good performance. Meanwhile, the very small differences between the portable NIRS and benchtop NIRS could be found. This further proved that they have the similar ability to predict the fatty acids of peanut samples.

Figure 2.5 showed the best prediction model results for the major fatty acids in HOP based on portable NIRS and benchtop NIRS. Analysing this figure, it was apparent that both the calibration data (represented by the blue circles) and the validation data (represented by the red triangles) were in good agreement with the resulting model. On the other hand, for normal peanuts, the best pretreatment for oleic acid, linoleic acid, and palmitic acid based on portable NIRS was FD+DT, DT+SD, and SD+DT respectively. Compared with portable NIRS, the best pretreatment for oleic acid, linoleic acid, and palmitic acid for benchtop NIRS was FD+SVN, DT+FD, and SD+DT individually. The calibration and validation models presented similar performance for both types of equipment. For example, the best results of oleic acid based on the portable NIRS were that R_C and R_P with their corresponding SEC and SEP were 0.90 and 0.87, 1.31 and 1.10 respectively. On the other hand, the best models according to the benchtop NIRS were that R_C and R_P with their corresponding SEC and SEP were 0.93 and 0.87, 0.98 and 1.10 individually. From the above results, although portable NIRS obtained relatively little spectral information because of the spectral resolution, the spectral information related to fatty acids is sufficiently col-

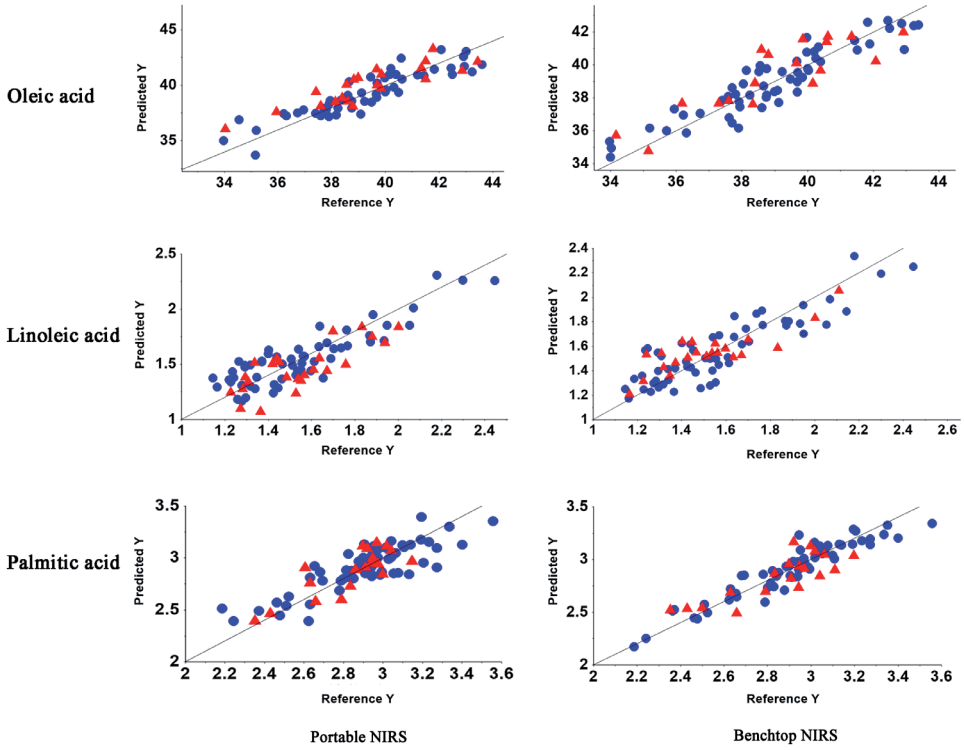


Figure 2.5. Best model results for the major fatty acids in high oleic acid peanuts by portable NIRS and benchtop NIRS (the blue circles represent the calibration model and the red triangles represent the validation data). (NIRS) Near-infrared spectroscopy.

lected. Therefore, it is obvious that both portable and benchtop NIRS shows good performance regarding the prediction of fatty acid contents in HOP and RP.

2.4. CONCLUSIONS

The portable NIRS and benchtop NIRS data combined with FD and SNV data pretreatment were subjected to PCA, which showed that HOP and RP were relatively well separated. PLS-DA showed that the accuracy for distinguishing the HOP and RP was 100% for both devices. For HOP, The R of calibration model of benchtop NIRS was 0.94, 0.90, and 0.94 with a SEC of 0.77, 0.12, and 0.10, respectively. Similarly, the R of the calibration model of the portable NIRS was 0.90, 0.88, and 0.88 for oleic acid, linoleic acid, and palmitic acid with a SEC of 0.97, 0.12, and 0.12, respectively. The RPD of all models were over 2 which showed good performance of the models. Therefore, the portable NIRS devices combined with chemometrics can replace the benchtop NIRS to classify the HOP and predict the major fatty acids, which can reduce equipment costs and allow in-situ measurements. That latter can make raw material purchase and breed varieties faster and more efficient.

REFERENCES

Agyekum, A. A., Kutsanedzie, F. Y. H., Mintah, B. K., Annavaram, V., Zareef, M., Hassan, M. M., Arslan, M., & Chen, Q. (2019). Rapid and Nondestructive Quantification of Trimethylamine by FT-NIR Coupled with Chemometric Techniques. *Food Analytical Methods, 12*(9), 2035-2044. https://doi.org/10.1007/s12161-019-01537-0.

Cascant, M. M., Breil, C., Fabiano-Tixier, A. S., Chemat, F., Garrigues, S., & de la Guardia, M. (2018). Determination of fatty acids and lipid classes in salmon oil by near infrared spectroscopy. *Food Chemistry, 239*, 865-871. https://doi.org/10.1016/j.foodchem.2017.06.158.

Cirilli, M., Bellincontro, A., Urbani, S., Servili, M., Esposto, S., Mencarelli, F., & Muleo, R. (2016). On-field monitoring of fruit ripening evolution and quality parameters in olive mutants using a portable NIR-AOTF device. *Food Chemistry, 199*, 96-104. https://doi.org/10.1016/j.foodchem.2015.11.129.

de Lima, G. F., Andrade, S. A. C., da Silva, V. H., & Honorato, F. A. (2018). Multivariate Classification of UHT Milk as to the Presence of Lactose Using Benchtop and Portable NIR Spectrometers. *Food Analytical Methods, 11*(10), 2699-2706. https://doi.org/10.1007/s12161-018-1253-7.

Dettori, M., Arru, B., Azara, A., Piana, A., Mariotti, G., Camerada, M. V., Stefanelli, P., Rezza, G., & Castiglia, P. (2018). In the Digital Era, Is Community Outrage a Feasible Proxy Indicator of Emotional Epidemiology? The Case of Meningococcal Disease in Sardinia, Italy. *International Journal of Environmental Research and Public Health, 15*(7), 1512. https://doi.org/10.3390/ijerph15071512.

Egesel, C. Ö., Kahrıman, F., Ekinci, N., Kavdır, İ., & Büyükcan, M. B. (2016). Analysis of Fatty Acids in Kernel, Flour, and Oil Samples of Maize by NIR Spectroscopy Using Conventional Regression Methods. *Cereal Chemistry Journal, 93*(5), 487-492. https://doi.org/10.1094/cchem-12-15-0247-r.

Feng, L., Zhang, M., Adhikari, B., & Guo, Z. (2019). Nondestructive Detection of Postharvest Quality of Cherry Tomatoes Using a Portable NIR Spectrometer and Chemometric Algorithms. *Food Analytical Methods, 12*(4), 914-925. https://doi.org/10.1007/s12161-018-01429-9.

García Martín, J. F. (2015). Optical path length and wavelength selection using Vis/NIR spectroscopy for olive oil's free acidity determination. *International Journal of Food Science & Technology, 50*(6), 1461-1467. https://doi.org/10.1111/ijfs.12790.

Gong, A., Shi, A., Liu, H., Yu, H., Liu, L., Lin, W., & Wang, Q. (2018). Relationship of chemical properties of different peanut varieties to peanut butter storage stability. *Journal of Integrative Agriculture, 17*(5), 1003-1010. https://doi.org/10.1016/s2095-3119(18)61919-7.

Gromski, P. S., Muhamadali, H., Ellis, D. I., Xu, Y., Correa, E., Turner, M. L., & Goodacre, R. (2015). A tutorial review: Metabolomics and partial least squares-discriminant analysis--a marriage of convenience or a shotgun wedding. *Analytica Chimica Acta, 879*, 10-23. https://doi.org/10.1016/j.aca.2015.02.012.

Lin, C., Chen, X., Jian, L., Shi, C., Jin, X., & Zhang, G. (2014). Determination of grain protein content by near-infrared spectrometry and multivariate calibration in barley. *Food Chemistry, 162*, 10-15. https://doi.org/10.1016/j.foodchem.2014.04.056.

Liu, N., Parra, H. A., Pustjens, A., Hettinga, K., Mongondry, P., & van Ruth, S. M. (2018). Evaluation of portable near-infrared spectroscopy for organic milk authentication. *Talanta, 184*, 128-135. https://doi.org/10.1016/j.talanta.2018.02.097.

Liu, S., Li, H., Hassan, M. M., Zhu, J., Wang, A., Ouyang, Q., Zareef, M., & Chen, Q. (2019). Amplification of Raman spectra by gold nanorods combined with chemometrics for rapid classification of four Pseudomonas. *International Journal of Food Microbiology, 304*, 58-67. https://doi.org/10.1016/j.ijfoodmicro.2019.05.020.

Ma, Y. B., Babu, K. S., & Amamcharla, J. K. (2019). Prediction of total protein and intact casein in cheddar cheese using a low-cost handheld short-wave near-infrared spectrometer. *Lwt, 109*, 319-326. https://doi.org/10.1016/j.lwt.2019.04.039.

Malegori, C., Nascimento Marques, E. J., de Freitas, S. T., Pimentel, M. F., Pasquini, C., & Casiraghi, E. (2017). Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. *Talanta, 165*, 112-116. https://doi.org/10.1016/j.talanta.2016.12.035.

Nawade, B., Mishra, G. P., Radhakrishnan, T., Dodia, S. M., Ahmad, S., Kumar, A., Kumar, A., & Kundu, R. (2018). High oleic peanut breeding: Achievements, perspectives, and prospects. *Trends in Food Science & Technology, 78*, 107-119. https://doi.org/10.1016/j.tifs.2018.05.022.

Ncama, K., Opara, U. L., Tesfay, S. Z., Fawole, O. A., & Magwaza, L. S. (2017). Application of Vis/NIR spectroscopy for predicting sweetness and flavour parameters of 'Valencia' orange (Citrus sinensis) and 'Star Ruby' grapefruit (Citrus x paradisi Macfad). *Journal of Food Engineering, 193*, 86-94. https://doi.org/10.1016/j.jfoodeng.2016.08.015.

Nepote, V., Mestrallet, M. G., Accietto, R. H., Galizzi, M., & Grosso, N. R. (2006). Chemical and sensory stability of roasted high-oleic peanuts from Argentina. *Journal of the Science of Food and Agriculture, 86*(6), 944-952. https://doi.org/10.1002/jsfa.2442.

Olmedo, R., Ribotta, P., & Grosso, N. R. (2018). Oxidative stability, affective and discriminative sensory test of high oleic and regular peanut oil with addition of oregano essential oil. *Journal of Food Science and Technology, 55*(12), 5133-5141. https://doi.org/10.1007/s13197-018-3459-5.

Qiang, W. (2018). Quality characteristics and determination methods of peanut raw materials. In *Peanut processing characteristics and quality evaluation* (pp. 69-125): Springer.

Quintelas, C., Mesquita, D. P., Ferreira, E. C., & Amaral, A. L. (2019). Quantification of pharmaceutical compounds in wastewater samples by near infrared spectroscopy (NIR). *Talanta, 194*, 507-513. https://doi.org/10.1016/j.talanta.2018.10.076.

Robinson, J., Singh, R., & Kays, S. (2008). Evaluation of an automated hydrolysis and extraction method for quantification of total fat, lipid classes and trans fat in cereal products. *Food Chemistry, 107*(3), 1144-1150. https://doi.org/10.1016/j.foodchem.2007.09.041.

Sheng, R., Cheng, W., Li, H., Ali, S., Akomeah Agyekum, A., & Chen, Q. (2019). Model development for soluble solids and lycopene contents of cherry tomato at different temperatures using near-infrared spectroscopy. *Postharvest Biology and Technology, 156*. https://doi.org/10.1016/j.postharvbio.2019.110952.

Soto-Barajas, M. C., Gonzalez-Martin, M. I., Salvador-Esteban, J., Hernandez-Hierro, J. M., Moreno-Rodilla, V., Vivar-Quintana, A. M., Revilla, I., Ortega, I. L., Moron-Sancho, R., & Curto-Diego, B. (2013). Prediction of the type of milk and degree of ripening in cheeses by means of artificial neural networks with data concerning fatty acids and near infrared spectroscopy. *Talanta, 116*, 50-55. https://doi.org/10.1016/j.talanta.2013.04.043.

Suchoszek-Łukaniuk, K., Jaromin, A., Korycińska, M., & Kozubek, A. (2011). Chapter 103 – Health Benefits of Peanut (Arachis hypogaea L.) Seeds and Peanut Oil Consumption. *Nuts & Seeds in Health & Disease Prevention*, 873-880.

Tillman, B. L., Gorbet, D. W., & Person, G. (2006). Predicting oleic and linoleic acid content of single peanut seeds using near-infrared reflectance spectroscopy. *Crop Science, 46*(5), 2121-2126.

Vassiliou, E. K., Gonzalez, A., Garcia, C., Tadros, J. H., Chakraborty, G., & Toney, J. H. (2009). Oleic acid and peanut oil high in oleic acid reverse the inhibitory effect of insulin production of the inflammatory cytokine TNF-alpha both in vitro and in vivo systems. *Lipids in Health and Disease, 8*, 25. https://doi.org/10.1186/1476-511X-8-25.

Wang, J., Zareef, M., He, P., Sun, H., Chen, Q., Li, H., Ouyang, Q., Guo, Z., Zhang, Z., & Xu, D. (2019). Evaluation of matcha tea quality index using portable NIR spectroscopy coupled with chemometric algorithms. *Journal of the Science of Food and Agriculture, 99*(11), 5019-5027. https://doi.org/10.1002/jsfa.9743.

Wu, D., Chen, X., Shi, P., Wang, S., Feng, F., & He, Y. (2009). Determination of alpha-linolenic acid and linoleic acid in edible oils using near-infrared spectroscopy improved by wavelet transform and uninformative variable elimination. *Analytica Chimica Acta, 634*(2), 166-171. https://doi.org/10.1016/j.aca.2008.12.024.

Wu, D., & Sun, D. W. (2013). Potential of time series-hyperspectral imaging (TS-HSI) for non-invasive determination of microbial spoilage of salmon flesh. *Talanta, 111*, 39-46. https://doi.org/10.1016/j.talanta.2013.03.041.

Xiaobo, Z., Xiaowei, H., & Povey, M. (2016). Non-invasive sensing for food reassurance. *Analyst, 141*(5), 1587-1610. https://doi.org/10.1039/c5an02152a.

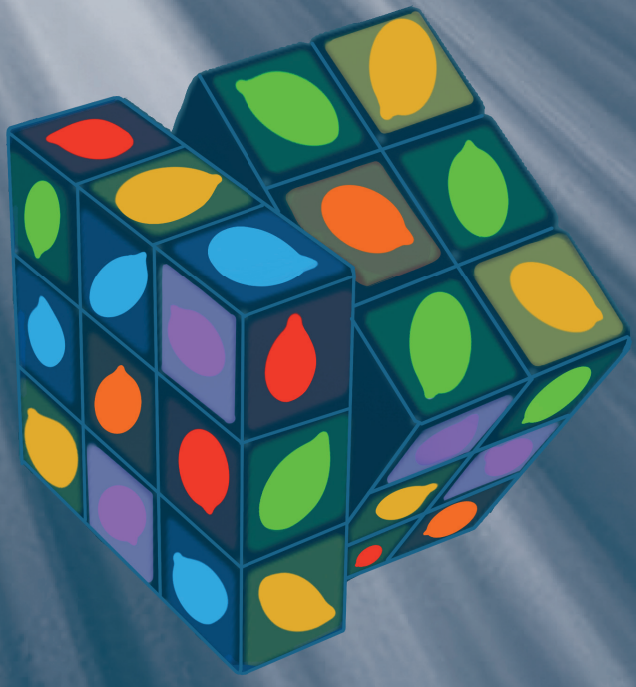
Yan, J., van Stuijvenberg, L., & van Ruth, S. M. (2019). Handheld Near-Infrared spectroscopy for distinction of extra virgin olive oil from other olive oil grades substantiated by compositional data. *European Journal of Lipid Science and Technology, 121*(12), 1900031. https://doi.org/10.1002/ejlt.201900031.

Yang, M., Chen, Q., Kutsanedzie, F. Y. H., Yang, X., Guo, Z., & Ouyang, Q. (2017). Portable spectroscopy system determination of acid value in peanut oil based on variables selection algorithms. *Measurement, 103*, 179-185. https://doi.org/10.1016/j.measurement.2017.02.037.

Ye, M., Gao, Z., Li, Z., Yuan, Y., & Yue, T. (2016). Rapid detection of volatile compounds in apple wines using FT-NIR spectroscopy. *Food Chemistry, 190*, 701-708. https://doi.org/10.1016/j.foodchem.2015.05.112.

Yu, H., Liu, H., Wang, N., Yang, Y., Shi, A., Liu, L., Hu, H., Mzimbiri, R. I., & Wang, Q. (2016). Rapid and visual measurement of fat content in peanuts by using the hyperspectral imaging technique with chemometrics. *Analytical Methods, 8*(41), 7482-7492. https://doi.org/10.1039/c6ay02029a.

Yu, H., Wang, Q., Shi, A. M., Yang, Y., Liu, L., Hu, H., & Liu, H. Z. (2017). Visualization of Protein in Peanut Using Hyperspectral Image with Chemometrics. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis, 37*(3), 853-858. https://doi.org/10.3964/j.issn.1000-0593(2017)03-0853-06.



CHAPTER 3

Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics

This chapter has been published as:
Yu, H., Liu, H., Erasmus, S. W., Zhao S., Wang, Q., & van Ruth, S. M. (2020). Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics. Industrial Crops and Products, 158, 112956.

ABSTRACT

The quality traits of peanuts (*Arachis hypogaea L.*) are fundamental to the whole peanut industry. However, many common analyses require the samples to be brought to the laboratory. Therefore, this research explores the feasibility of portable near-infrared spectroscopy combined with a single detection accessory to analyse the composition of peanuts in a single seed level quantitatively. The single detection accessory was specifically designed for spectral data collection considering the internal and external characteristics of single peanuts. Confocal laser scanning microscopy revealed that the oil body and protein body were randomly distributed at cell of single peanuts. The external characteristics of single peanuts were also determined and considered length (11.32-24.25 mm) and width (7.49-12.25 mm). The chemical compositional data (i.e. fat, sucrose, protein, and 16 amino acids) were determined by conventional wet-chemical methods and showed large variation. Principal component analysis on the compositional data showed that peanuts with higher fat contents usually have higher hydrophobic amino acids contents, lower sucrose contents, and lower protein contents. The composition prediction models of single peanuts were estimated using partial least squares regression models that were integrated with different spectral pre-treatments and validated by external sets. The results showed that the prediction models have good performance with a correlation coefficient above 0.88 (calibration) and 0.83 (prediction) and a residual prediction deviation above 1.5 except for a few indicators. Overall, portable near-infrared spectroscopy offered reliable methods to assess the major components and amino acids quantitatively in a single peanut, which will improve the raw material quality in the peanut industry through the simultaneous and short-term determination of multiple indicators.

Keywords: Amino acids; Partial least squares regression; Peanut; Portable NIRS device; Single kernels; Sucrose

3.1. INTRODUCTION

Peanut (*Arachis hypogaea L.*) is a widespread leguminous crop and a dominant commercial agricultural product. Global peanut production was over 48 million tons (Mt) in 2017. China (17.2 Mt), the world's largest peanut-producing country, accounted for nearly 36% of the total production (Faostat, 2020). Peanuts are nutritious and mainly comprise fat, protein, sucrose, and amino acids. (Wang, 2016, 2018). The fat component accounts for half of peanuts' contents and are the prevailing nutrient constituent in peanuts for producing edible oil (Yu et al., 2016). Proteins are the second group of constituents in peanuts and consist of eight essential amino acids which satisfy the demands of the Food and Agriculture Organization, except for methionine (Dubinina et al., 2014). The non-essential amino acids not only carry health effects but also have great contributions to the function of proteins such as their soluble and gel characteristics which depend on the molecular structure of amino acids (Wang, 2016; Yu et al., 2017). The sweet and attractive flavour of peanuts and peanut products are derived from their sucrose as well as Maillard reactions with amino acids (Wang et al., 2017). There is no doubt that peanuts consisting of high levels of fat, protein, and sucrose are preferred over others by the peanut industry. Breeding experts, seed retailers, peasants, food and feed companies, and other stakeholders all benefit from high-quality peanut varieties. Hence, it is an ongoing aim of the peanut industry to cultivate new peanut varieties with specific traits.

It is extremely useful for breeders and other stakeholders in the peanut supply chain to master the genetic attributes of peanut qualities and to select desired peanuts by analysing the quality traits in single kernels. The conventional methods for screening peanuts carried out in laboratories are time-consuming, complex operations, and usually result in sample destruction. For breeders and others, sometimes a limited sample size is available and one may not have a lot of materials for evaluation by conventional methods. Therefore, it is essential to explore techniques that are speedy, cost-effective, and non-damaged in nature for the detection of single peanut kernels quality. Consequently, near infrared spectroscopy (NIRS) has great advantages of multiple-trait, non-destructive analysis in a relatively short time (Agelet et al., 2012; Chopra et al., 2019).

Normally, NIRS is bulk sample analyser that requires, on average, samples sizes of about 250 g peanut kernels (Agelet & Hurburgh, 2014). The measurements of fat (Sundaram et al., 2010), oleic acid (Shin et al., 2010), protein, and amino acids (Wang et al., 2013) have been conducted by NIRS which proved its potential applications in peanuts. However, research on single peanut kernels only focused on fatty acid determinations (Tillman et al., 2006). By contrast, NIRS of single seeds was also already applied for grains and beans analysis. Specifically, the contents of amino acids and sucrose (Natsuga et al., 2007) of single soybean kernels have been examined. Furthermore, the protein, starch

(Jiang et al., 2007), and ergosterol contents (Berardo et al., 2005) of single corn kernels, glucosinolates and indole concentrations of single rapeseed kernels (Hom et al., 2006), as well as caffeine concentrations of single coffee kernels (Fox et al., 2013) have been analysed by NIRS. Meanwhile, with the miniaturisation of optical technology, portable NIRS has received more and more attention because of its lower price and higher convenience than the benchtop equivalents, at least for some applications. Although the studies mentioned above adopted benchtop NIRS, some studies have been conducted with portable NIRS and used the technology for quantitative testing, such as for the fatty acid analysis of peanuts (Yu et al., 2020), and for qualitative testing such as organic milk distinction (Liu et al., 2018) and olive oil classification (Yan et al., 2019). Satisfactory conclusions of the above research were made by portable NIRS coupled with chemometrics.

To analyse single peanut kernels, a detection accessory combined with portable NIRS should be designed with consideration of the characteristics of single peanut kernels. However, no reference study about features of single peanut kernels to help design such a detection accessory in order to offer multiple-trait non-destructive analysis was published so far. Therefore, this study aims to investigate the external and internal characteristics of single peanut kernels for designing a single peanut detection accessory for spectral data collection. Furthermore, the results obtained with the portable NIRS device was examined for its prediction capabilities of the fat, protein, sucrose, and amino acids contents of single peanuts. The relationships between spectral data and the different molecular groups of those components were also explored.

3.2. MATERIALS AND METHODS

3.2.1. Peanut Samples

A set of 110 peanut varieties and strains were randomly selected from 10 main planting provinces in China. Peanuts (approximately 2 kg per variety and strain) were stored at 4°C in a commercial cold store (Yuandong Co., Ltd., Tianjin, China). The overview of all peanut varieties and strains collected and analyses are provided as Supplementary data (Appendix A. Supplementary data can be found in the online version, at https://doi.org/10.1016/j.indcrop.2020.112956).

3.2.2. Compositional analysis

The following number of samples were randomly analysed per conventional compositional analysis: 110 (fat and protein), 100 (amino acids), and 80 (sucrose). More information on the sample/analysis distribution can be found in Supplementary data (Appendix

A. Supplementary data can be found in the online version, at https://doi.org/10.1016/j.indcrop.2020.112956). Each chemical analysis was performed in duplicate and the results were averaged for data analysis. The reference fat contents were determined based on Soxhlet extraction method (GB5009.6, 2016) by using a Soxtec 2050 instrument (FOSS, Hillerød, Denmark). The protein contents were measured by Kjeldahl method (GB5009.5, 2016) using a 2300 Nitrogen Analyser (FOSS, Hillerød, Denmark) with a conversion coefficient (5.46). The quantitative analyses of amino acids were conducted through the standard (GB5009.124, 2016). Briefly, about 2 g peanut samples were transferred to hydrolysis tubes with nitrogen after acidic hydrolysis with 15 mL hydrochloric acid solution. The hydrolysis tubes were placed in a hydrolysis furnace (110°C) for 22 h. After constant volume and drying, the solution was then filtered through a 0.22 μm membrane after which the filtrate was analysed by an L-8900 amino acids automatic analyser (Hitachi, Tokyo, Japan). Sucrose in the peanuts was analysed by high-performance anion-exchange chromatography method (GB5009.8, 2016). Approximately 5 g peanut powder samples were degreased firstly and then extracted in volumetric flasks (100 mL) placed in an ultrasonic bath for 30 min with water, zinc acetate solution, and potassium ferrocyanide solution. The extracts were centrifuged for 30 min at 12,000 rpm. The supernatant was collected and filtered through a 0.45 μm filter after which the filtrate was analysed using a Dionex ICS-3000 (Thermo Fisher Scientific, Waltham, the USA). All reagents used in the above analyses were purchased from Sinopharm Chemical Reagent Co., Ltd. (Beijing, China).

3.2.3. Size analysis of single peanut kernels

The width and length of single peanut kernels were measured using a Vernier Caliper (Shanghai Tool Factory Co., Ltd., Shanghai, China). The width (mm) of peanut kernels is the widest point of the mature plump seeds, whereas the longest length of the mature plump seeds is regarded as the length (mm) of peanut kernels according to a previously reported procedure (Jiang et al., 2006). All varieties and strains (n = 110) were analysed in duplicate.

3.2.4. Confocal laser scanning microscopy analysis

The internal microstructures of the peanut kernels were assessed by confocal laser scanning microscopy (CLSM) using a Zeiss LSM 880 microscope (Carl Zeiss Jena, Munchen, Germany). The microscope was outfitted with 40×objectives and two lasers, including an Argon laser (λ excitation 488 nm) and a diode-pumped-solid-state laser (λ excitation 561 nm). Matured embryos were separated from peanut seeds based on an approach reported previously (Perry & Wang, 2003). In brief, peanut embryos were immediately soaked for 4 h in 2.5% v/v glutaraldehyde and 1.6% w/v paraformaldehyde in a 0.1 M phosphate buffer solution (pH 7.4). After flushing, the embryos were soaked in a 2% w/v osmium tetroxide solution for an additional 4 h followed by dehydration using acetone. The dehydrated

embryos were fixed in epoxy resin and semithin sections were attained by a ultramicrotome EM UC6 (Leica, Wetzlar, Germany). All reagents used in the above processing of semithin sections were purchased from Sinopharm Chemical Reagent Co., Ltd. (Beijing, China). 0.1% w/v Nile red and 0.1% w/v fluorescein isothiocyanate (Sigma Chemical Co., St Louis, MO, the USA) were used for visualization of the fat and protein components. The samples Yuanza9847, Hanghua2, Yueyou7, Jihua13, Yuhua9326, and Huayu25 were selected as representative samples for analysis and were analysed in duplicate.

3.2.5. Portable NIRS and spectral collection

A previously developed box-portable NIRS system was used in this study (Yu et al., 2020). The device comprised a spectrometer-Micro-NIRS 1700 (908-1676 nm, 6nm interval, VIAVI Solutions, San Jose, the USA), a control centre consisting of a tablet computer (Surface, Microsoft Corporation, Redmond, the USA), a reference white Teflon board for white reference, a lithium battery for power supply, and a single peanut kernels detection accessory designed in this study. Each peanut sample (n = 110) was placed into the single detection accessory to collect the spectral data by measurement in reflection mode. The dark reference was regarded as the collected dark current when the light was off. The measurements (spectral data acquisitions) were conducted after dark and white reference calibration. Twenty peanut kernels per sample were scanned once. The spectral data per sample were averaged for data analysis.

3.2.6. Data analysis

The first step was to determine outliers. The average spectra and the corresponding components of single peanut kernels were regarded as the dependent variables X and Y, respectively. The X outliers of each component group were determined by principal component analysis (PCA) according to the Hotelling's T^2 (Yu et al., 2020) separately, whilst the Y outliers were analysed by boxplot combined with violin analysis (Dettori et al., 2018; Xu et al., 2019). The boxplot analysis presented box charts with the median as the middle, the 25th and 75th percentiles as the edges, and the maximum and minimum value as the whiskers (Quintelas et al., 2019), while violin plots showed the distributions of data. The data outside the range of whiskers were considered the Y outliers which were marked with black dots. PCA combined with K-means clustering was used for exploring the relationship between different peanut varieties and strains. After the outliers were eliminated, samples of each component were first sorted by content in descending order. Fifteen samples (ten samples for sucrose) were selected for validation according to the sequence (e.g. select one from every six samples) and the remaining for calibration. The range and standard deviation (SD) of calibration and validation set of each component were calculated.

The partial least squares regression (PLSR) prediction models were established to quantify the composition of single peanut kernels (Chen et al., 2017; Kutsanedzie et al., 2018). Before modelling, the spectral data were pre-processed in order to eliminate the overlapping of the original spectra (Leng et al., 2020), such as the first derivative (1st der), the second derivative (2nd der), standard normal variate (SNV), detrending (DT), baseline (BL), multiple scattering correction (MSC), and combined pre-treatments including 1st der+DT, 2nd der+SNV, etc. After modelling, the performances of models were validated through cross validation and external prediction. The cross validation (leave-one-out) was used to avoid overfitting of the calibration models. The number of factors was chosen based on the best results of the cross validation. The evaluation index had the correlation coefficient in cross validation (R_{CV}) and standard error in cross validation (SECV). The external prediction, using the validation set, evaluated the robustness of the calibration models which included the parameters such as the correlation coefficient in calibration (R_C) and prediction (R_V), and standard error in calibration (SEC) and prediction (SEP) as given by Eq.(1), as well as bias and residual predictive deviation (RPD). The RPD is equal to the SD of validation set divided by SEP. Different RPD ranges have been reported in different sources (Fearn, 2002; Nicolaï et al., 2007). In this study, according to Lima (Lima et al., 2020) and Ye (Ye et al., 2016) research, if RPD values ≥ 2, models are considered excellent and RPD values < 1.5 are considered non-reliable, while the models with RPD value in between are considered good. R 3.5.3 (R Foundation for Statistical Computing, Austria) and Unscrambler 10.3 (CAMO Software AS, Norway) were used for data analysis.

$$\mathrm{SEP} = \sqrt{\frac{\sum_{i=1}^{n}(y_{pi}-\hat{y}_{pi})^2}{n-1}} \quad (1)$$

$\hat{y}_{pi}$ and y_{pi} are prediction and reference values of the *i*th sample in the validation group; *n* is the number of samples in the validation group. If all variables come from the calibration group, the equation indicates SEC.

3.3. RESULTS AND DISCUSSION

3.3.1. Compositional results of different peanut varieties

The distribution map composed of a boxplot (black box in the violin) and violin plot shows the distributions of major components and amino acids (**Figure 3.1**). The fat content ranged from 43.5 to 57.6 g/100 g (with only one outlier identified), while the protein content ranged from 18.4 to 28.6 g/100 g (two outliers identified). Based on the violin plot, most fat and protein contents of peanut varieties ranged from 49-54 g/100 g and 22-27 g/100 g, respectively. This is similar compared with previous studies for fat (48-55 g/100 g) (Li et al., 2018)and protein (20-26 g/100 g) (Gong et al., 2018) in peanuts. Sucrose varied

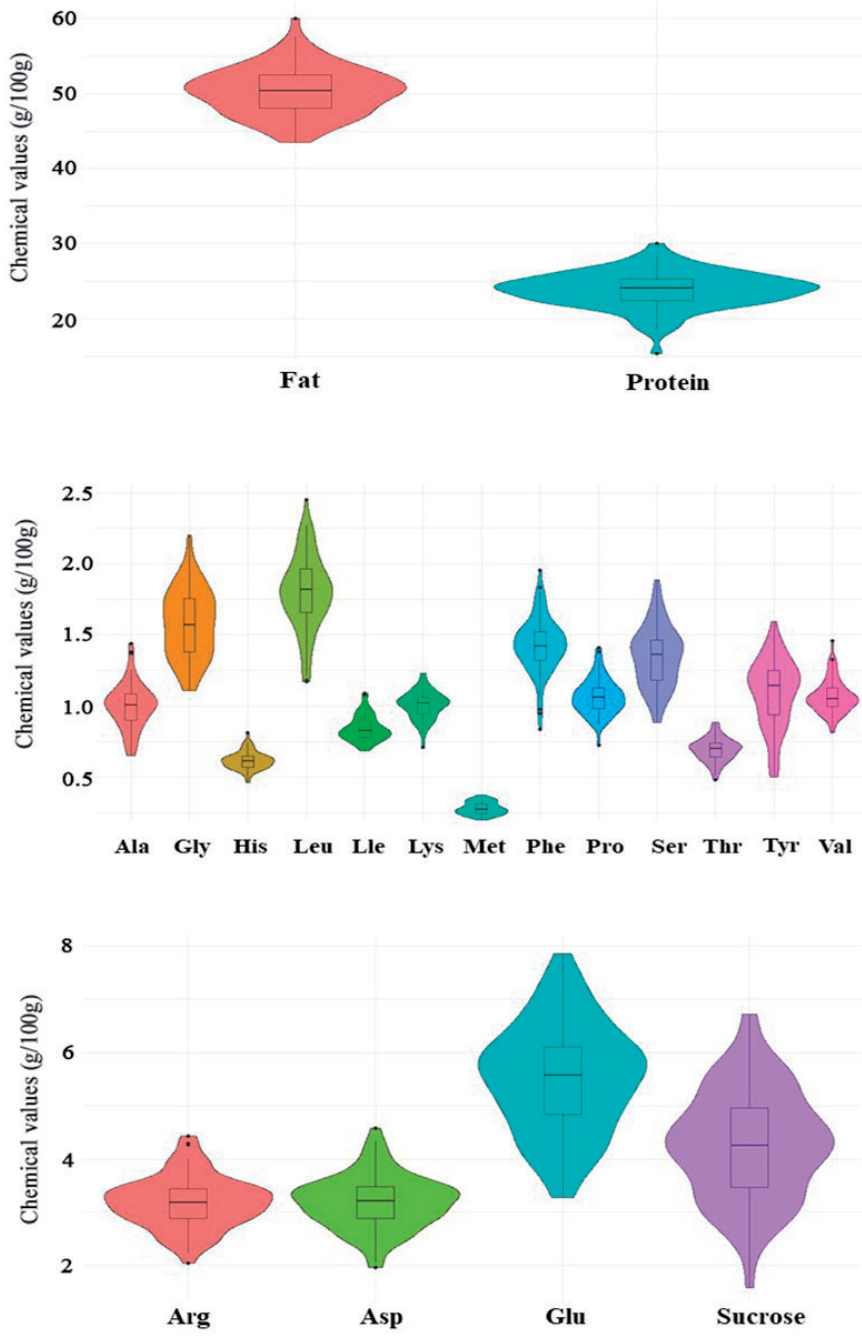


Figure 3.1. The distribution map of the different components of peanut kernels after deleting spectral outliers. (outliers marked with black spots) (Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (Phe) Phenylalanine; (Pro) Proline; (Ser) Serine; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

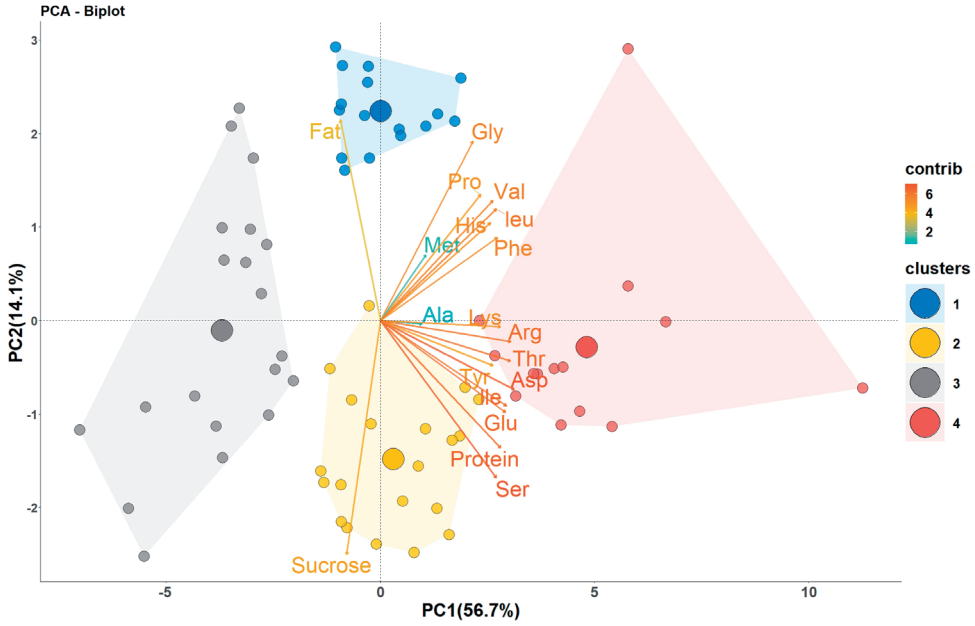


Figure 3.2. The principal component analysis (PCA) biplot of different peanut samples (n = 70, different coloured ovals indicate the different sample clusters and arrows indicate the contributions of compositions). (Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (Phe) Phenylalanine; (Pro) Proline; (Ser) Serine; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

from 1.6 to 6.7 g/100 g. Lower values (2.6-6.5 g/100 g) were observed by Bishi (Bishi et al., 2015). The peanuts appeared rich in essential amino acids. Especially, phenylalanine (Phe, 0.98-1.83 g/100 g, four outliers), leucine (Leu, 1.18-2.27 g/100 g, two outliers), and valine (Val, 0.81-1.31 g/100 g, two outliers) were main essential amino acids in the peanuts, while the other essential amino acids such as lysine (Lys, 0.71-1.23 g/100 g, one outlier), isoleucine (Ile, 0.68-1.09 g/100 g, an outlier), methionine (Met, 0.20-0.37 g/100 g, one outlier), and threonine (Thr, 0.50-0.88 g/100 g, an outlier) were relatively low. The non-essential amino acids presenting the highest concentrations were glutamic acid (Glu, 3.27-7.86 g/100 g), aspartic acid (Asp, 2.05-4.33 g/100 g, two outliers), arginine (Arg, 2.21-4.29 g/100 g, three outliers), and glycine (Gly, 1.11-2.20 g/100 g), whereas alanine (Ala, 0.65-1.37 g/100 g, two outliers), tyrosine (Tyr, 0.50-1.59 g/100 g), proline (Pro, 0.87-1.34 g/100g, three outliers), serine (Ser, 0.89-1.88 g/100 g), and histidine (His, 0.46-0.75 g/100 g, one outlier) displayed the lower concentrations. The amino acids contents were similar to those reported by Radhakrishnan (Radhakrishnan et al., 2014) and Klevorn (Klevorn et al., 2019).

The PCA method was used to explore the relationship between different peanut samples (n = 70) containing with all components analysed. The PCA biplot (PCA score plot

combined with loading plot) is shown in the **Figure 3.2**. The total number of PC1 and PC2 is 70.3%. K-means as an unsupervised method was used to cluster the peanuts based on all components analysed. The blue points present that samples with a higher fat content, while the yellow ones have higher sucrose contents. Obviously, a negative correlation between fat and sucrose contents exist, which is consistent with previous research (Bishi et al., 2015). It is also found that fat and protein contents are negatively correlated because they are located on opposite sides of the plot, which another study confirmed as well (Sarvamangala et al., 2011). The amino acids have the same direction, which means they are positively related with the fat content. Among them, hydrophobic amino acids such as Leu, Met, Val, and Phe have similar orientation to fat. It appears that peanut varieties and strains with higher fat contents usually have higher hydrophobic amino acids contents, lower sucrose contents, and lower protein contents.

3.3.2. The internal structure of single peanuts and compositional distribution

The distribution of the oil body (OB) and protein body (PB) which are the source of NIRS information in the peanut cell was investigated by CLSM. **Figure 3.3** shows that OB (red) and PB (green) were randomly and evenly (no agglomeration) distributed at cell level, which is consistent with previous results (Zaaboul et al., 2018). The amino acids, especially the essential ones, are known to be located in the OB and PB. Therefore, the collected spectral information would likely contain fat, protein and amino acids information at the same time. This is one of the reasons why spectral pre-processing are performed before modelling. There are a lot of differences between different varieties and strains. For example, the differences between Yuanza9807 (**Figure 3.3a**) and Jihua13 (**Figure 3.3b**) are considerable since most PB in Yuanza9807 are larger but the number is relatively lower because of the cell size and PB size. OB are very densely distributed in cells of two varieties and it could be observed that OB in Jihua13 are larger. The superposition of differences at cell level eventually leads to differences in chemical composition between varieties and strains.

3.3.3. Development of the single detection accessory

The range of distribution of the length and width of all peanuts (n = 110) was 11.32-24.25 mm and 7.49-12.25 mm, respectively. To meet different sizes, two different apertures were designed considering the peanut size range. One was sized 12.5 mm×24.5 mm and the other one was 9.7 mm×17.5 mm (**Figure 3.4a**). The bottom part of the accessory was made of 1 mm quartz glass which has good permeability for light in the NIRS region. The material used for the rest of the accessory was aluminium. According to previous research (Fraser, 2001), wavelengths from 1400 to 1600 nm could penetrate up to 1 mm. Although the thickness of the penetration would be increased as the wavelength

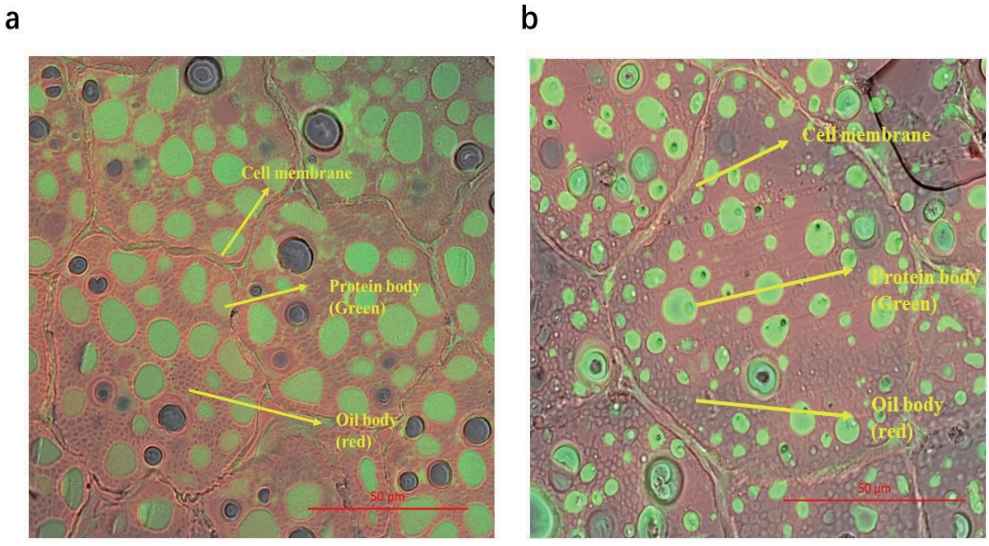


Figure 3.3. The confocal images of single peanut kernels, (a) Yuanza9847, (b) Jihua13.

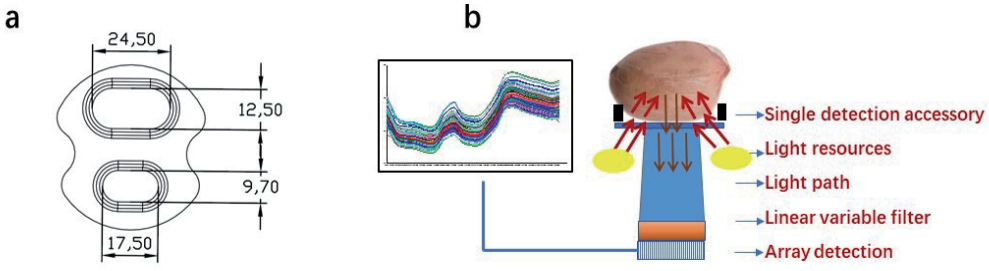


Figure 3.4. The (a) single kernel near-infrared spectroscopy detection accessory (the numerals represent the aperture size, unit: mm), and (b) schematic of the portable near-infrared spectroscopy measurement system.

goes down, it is still very difficult to penetrate peanuts ($width_{min}$ = 7.49 mm). Therefore, the reflectance measurements for peanut kernel detection were chosen. A schematic diagram of the portable device with a single peanut accessory and the spectral acquisition route are shown in the **Figure 3.4b**.

3.3.4. Spectral data and various data pre-treatment procedures

The original averaged spectral data of the single kernels of all peanut varieties (n = 110) are presented in the **Figure 3.5a**. Compared with previous research (Wang et al., 2013; Yu et al., 2016), the curve of the spectral line was the same. Specifically, the peak and valley of the original spectra appeared in the same position. The peaks of the original spectra are 1205 nm and 1460 nm, while 1106 nm and 1298 nm are the valleys, which are caused by the frequency absorption harmony of hydrogen-containing groups including

C-H (generally from fat and sucrose), O-H (generally from water), and N-H (generally from protein and amino acids) (Hourant et al., 2000). Hence, the peaks and valleys are characteristic of the main chemical constituents of the peanuts.

Different pre-treatments of the raw spectra could help to attain more useful information for the generation of models. For instance, derivatives could eliminate the baseline and the background effect and improve the resolution and sensitivity of spectral information (Chu, 2011). Detrending is a method to erase the baseline drift of the spectral data (Pérez-Rodríguez et al., 2018). The SNV or MSC pre-processing can remove the unnecessary spectral information induced by seed size (Agelet et al., 2012). After pre-treatment of the spectra, more signals of molecular groups that may be related to fat, sucrose, protein, and amino acids were exposed. Taking the 2nd derivate combined with detrending as an example (**Figure 3.5b**), 927 nm is the third C-H stretch overtone from methylene (Workman Jr & Weyer, 2012), while 952 nm is mainly attributed to the second O-H stretch overtone (Williams & Norris, 1987). There are many signals between 1110 nm and 1250 nm linked with the C-H stretch overtone. More specifically, It is found that 1125 nm and 1181 nm come from the second C-H stretch overtone (Fernandez-Novales et al., 2019) and the second C-H stretching of HC=CH respectively. It is reported that 1212 nm is derived from the second overtone C-H stretching of CH_2-CH_2 and 1243 nm is from the band assignments of 3×C-H stretch (Workman Jr et al., 2012). The first overtone region exhibits the major bonded O-H peak (1280 nm) whereas the C-H combination of C-H_3 is located at 1385 nm (Workman Jr et al., 2012). The wavelength 1428 nm is usually identified as the first stretching of N-H and the signal at 1658 nm is the C-H methyl of carbonyl adjacent (Workman Jr et al., 2012).

The PCA results of the original spectra of peanut varieties (n = 110) are indicated in **Figure 3.5c**. The total variation explained by PC1 and PC2 was 99%. It could be found that some dots are outside the line calculated the Hotelling's T^2 with 0.5% significance level. These dots are regarded as the outliers and there were five outliers in total. These five outliers also exist in the amino acids data set. After removing all chemical and spectral outliers, descriptive statistics including the range values and the SD of the chemical composition of the reference samples were calculated (**Table 3.1**). The wide-range data and the SD of different compositions of peanut kernels show the sample diversity in this study.

3.3.5. Prediction models for compositional characteristics

Various models were calculated to predict the compositional characteristics of the peanuts from the spectral data. The calibration, cross-validation, and external prediction statistics arguments (PLSR factor, R, SEC/CV/P, RPD, and bias) of the optimisation models for the assessment of the major compounds and each amino acid combined with the optimal pre-treatment are shown in **Table 3.2**. Generally, the number of PLSR factors is

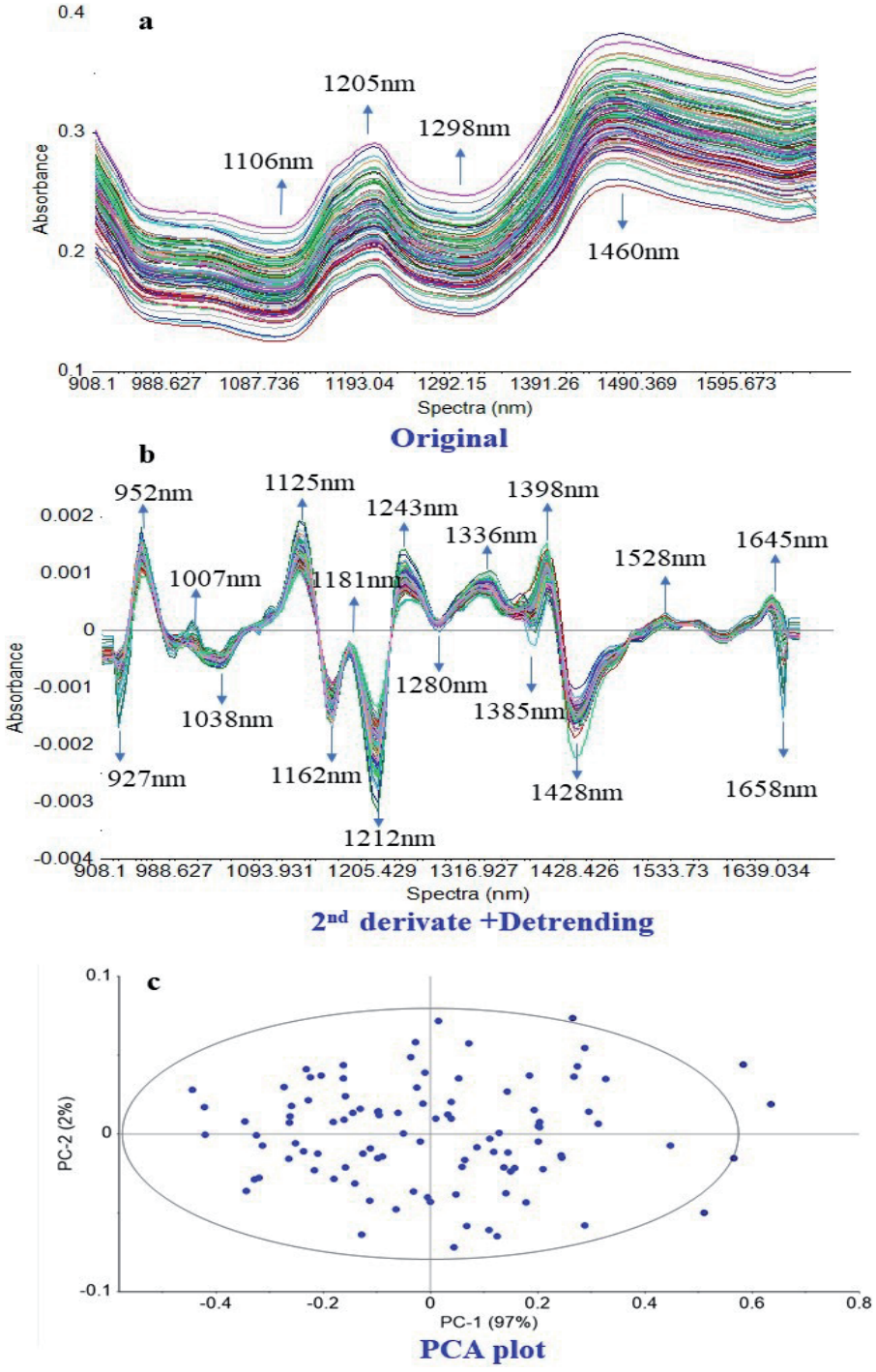


Figure 3.5. The (a) raw spectra and (b) pre-processed spectra (2nd derivate combined with detrending) of single peanut kernels, and the (c) principal component analysis (PCA) scores plot.

Table 3.1. Compositional characteristics of sample sets used for the calibration and external validation of the spectral models.

Component (g/100 g)	Calibration set			Validation set		
	n	Range	SD	n	Range	SD
Fat	89	43.5-57.6	3.1	15	45.9-56.7	4.2
Protein	88	18.4-28.6	2.2	15	18.4-26.8	3.4
Sucrose	70	1.6-6.7	1.0	10	2.4-6.0	1.2
Arg	77	2.21-4.29	0.45	15	2.27-4.40	0.49
Asp	78	2.05-4.33	0.47	15	2.58-4.25	0.47
Phe	76	0.98-1.83	0.17	15	1.00-1.72	0.19
Ala	78	0.65-1.37	0.15	15	0.74-1.27	0.15
Gly	80	1.11-2.20	0.24	15	1.36-2.05	0.20
Lys	79	0.71-1.23	0.10	15	0.90-1.08	0.05
Tyr	80	0.50-1.59	0.24	15	0.81-1.40	0.17
Leu	78	1.18-2.27	0.25	15	1.47-2.09	0.18
Pro	77	0.87-1.34	0.11	15	0.88-1.29	0.10
Ile	79	0.68-1.09	0.09	15	0.68-0.88	0.06
Val	78	0.81-1.31	0.11	15	0.92-1.21	0.10
Met	80	0.20-0.37	0.04	15	0.20-0.35	0.05
Glu	80	3.27-7.86	1.03	15	3.82-6.81	0.82
Ser	80	0.89-1.88	0.22	15	1.06-1.76	0.19
Thr	79	0.50-0.88	0.08	15	0.59-0.76	0.06
His	79	0.46-0.75	0.06	15	0.46-0.77	0.07

(Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (n) Number of samples; (Phe) Phenylalanine; (Pro) Proline; (Ser) Serine; (SD) Standard deviation between samples; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

a critical parameter to evaluate the capability of the calibration models. If the number of PLSR factors used is too small, useful information would be lost. In contrast, the use of an excessive large number of factors would result in overfitting of the model. The best number of factors to use is determined by the results of cross validation. The factors of all components in single peanuts ranged from 7 to 11. Compared with bulk sample analysis, the spectral data of single kernels are easily affected by kernel size even if pre-treatment has been considered to reduce that noise. Therefore, the more useful information based on higher factors should be used. A similar number of factors were needed as the previous studies, such as for single rice kernels (factor = 9) (Xu. et al., 2019), single coffee beans (factor = 8-10) (Fox et al., 2013), and single cocoa beans (factor = 8-15) (Caporaso et al., 2018) to ensure robustness of the models.

The optimal calibration model for fat prediction was established with an RPD value of 2.39 (R_p = 0.91). This value is higher than the RPD value of 2.13 obtained by Tallada (Tallada et al., 2009) for single maize kernel samples. The model accuracy to predict the content of sucrose was also high considering the RPD value of 2.02 (R_p = 0.92). Very few studies

have focused on the prediction of sucrose content before. It was only reported that the model for the prediction of sucrose in green soybeans, using a transmission model, had a modest result with a R_{cv} value of 0.86 (Natsuga et al., 2007). Nevertheless, sucrose is the main contributor to the sweetness of peanuts (Bishi et al., 2013). Therefore, a good sucrose model as presented in this study is certainly relevant for peanuts. The protein content of the single peanut kernels could be predicted best by the model based on the baseline combined with MSC with an RPD value of 2.36 (R_p = 0.91). The result was slighter better than the previously reported models for protein measurements in single soybeans (RPD_{max} = 2.33) which were built by three different instruments except for the Light Tube (RPD = 3.28) with high resolution (3 nm) (Agelet et al., 2012). It also had good agreement with the RPD value of 2.9 obtained by Fox (Fox et al., 2011) for single barley analysis.

In terms of amino acids, good performance calibration models were established for Arg, Asp, Phe, Ile, Glu, and Ser. Although Arg is not the predominant component among all amino acids of peanuts, it is a characteristics amino acid of peanuts because its content is higher in peanuts compared with those of other crops and grains. The optimal model was built for the prediction of Arg with the highest calibration and validation correlation coefficients (R_c = 0.90; R_{cv} = 0.87; R_p = 0.90) by using 1st der combined with baseline pre-treatment. The RPD value of 1.98 was slightly lower than for bulk peanut samples (RPD = 2.88) (Wang et al., 2013). Though the matrices were different, these correlation coefficients were higher than the ones studied by Carbas (Carbas et al., 2020) for common beans. There is no doubt that Asp is another characteristic amino acid of peanuts that can adjust the metabolism of the human brain and nervous system. Good quality prediction models were built with the current dataset, showing high correlation coefficients for calibration (R_c = 0.90) and validation (R_p = 0.89). The RPD value of 2.15 was slightly lower than bulk peanut samples (RPD = 2.56) (Wang et al., 2013) and the same as Juan reported (RPD = 2.22) (Fernandez-Novales et al., 2019). It is known that Phe is the essential amino acid for most people for the synthesis of important neurotransmitters and hormones. However, the content should be controlled for all phenylketonuria because they cannot metabolise Phe. The model was first established for the prediction of Phe in peanut kernels and had good robustness (R_c = 0.89; R_p = 0.93; RPD = 2.65). These results were better than the Phe prediction model for rice wine (Shen et al., 2010) and common beans (Carbas et al., 2020). The model of Ile, which is one of the essential amino acids, was also built for the first time. **Table 3.2** showed that the model had the best performance (R_c = 0.84; R_p = 0.88; RPD = 2.00) with the pre-treatment, 1st der combined with MSC. Glu constitutes the highest content of amino acids in peanuts and Ser can promote the metabolism of fats and fatty acids. More importantly, these two amino acids had great contributions to the function of peanut protein (Wang et al., 2013). The models generated for detecting Glu and Ser in single peanut kernels had great results (RPD_{Glu} = 1.99; RPD_{Ser} = 2.08) which were better than those in similar research on intact grape berries (RPD_{Glu} = 1.90; RPD_{Ser} = 2.07) (Shen et

al., 2010). The rest of the essential amino acids included Lys, Thr, Leu, and Val. The models of these essential amino acids also had relatively good performance (R_p = 0.78-0.83; RPD = 1.50-1.58), which were better than the previous studies. Among these, Val and Leu are hydrophobic amino acids, while Lys and Thr are hydrophilic amino acids. Both affect the characteristics of peanut proteins. The rest of the non-essential amino acids included Ala, Gly, Tyr, Pro, and His. Except for His (RPD = 1.40), models for these amino acids showed good performance (R_p = 0.8-0.87; RPD = 1.54-1.72).

Bias in this kind of models is identified as the average differences between the predictive content and the actual content of a compound. A positive bias is regarded that the model overestimated the composition of an individual compound, while a negative value means that it is underestimated. A good model should have a low bias (Shao et al., 2011). To eliminate the mean order of magnitude effect, bias/mean was used instead of bias (Vallone et al., 2019). As shown in **Table 3.2**, most established models were slightly over-estimating the contents of the amino acids and major components in single peanuts. All the models had low bias/mean (< 0.002) except for Ala, Tyr, Leu, and Met. Meanwhile, those four amino acids had relatively low RPD values. It was further proven that models have more reliable performances leading to high RPD and low bias/mean.

3.4. CONCLUSIONS

Portable NIRS technology coupled with a single detection accessory allowed precise measurements of the major components and amino acids of single peanut kernels. This study highlighted the potential of the proposed portable single kernel instrument to estimate the individual composition of single peanuts. Meanwhile, high-fat peanut varieties and strains usually have relatively low protein and sucrose contents. Ultimately, it allows a better understanding of variation in contents among peanut kernels within and across groups of samples. The methodology is faster and high-throughput compared with conventional methods that require more time and transition of the samples to the laboratories. This offers a great advantage that can be used to control the quality and integrity of peanuts from the beginning of the industrial chain.

Table 3.2. The calibration, cross validation and external validation results of peanut kernel major components and amino acids.

Components	Pre-treatment	Factor	Calibration		Cross validation		External validation				
			Rc	SEC	Rcv	SECV	Rv	SEP	Bias/ Mean	Slope	RPD
Fat	2nd der+MSC	11	0.90	1.38	0.78	2.09	0.91	1.76	0.011	0.917	2.39
Sucrose	2nd der+SNV	10	0.94	0.39	0.83	0.62	0.92	0.57	0.011	1.164	2.02
Protein	BL+MSC	11	0.90	1.25	0.80	1.76	0.91	1.42	0.016	0.831	2.36
Arg	1st der+BL	9	0.92	0.19	0.87	0.24	0.90	0.25	-0.009	1.097	1.98
Asp	DT+SNV	10	0.90	0.22	0.82	0.29	0.89	0.22	-0.002	0.697	2.15
Phe	2nd der+DT	8	0.89	0.09	0.81	0.11	0.93	0.07	0.006	0.797	2.65
Ala	2nd der+MSC	10	0.90	0.07	0.77	0.11	0.83	0.09	-0.033	0.885	1.67
Gly	2nd der+BL	9	0.88	0.12	0.77	0.16	0.85	0.12	-0.019	0.990	1.63
Lys	1st der+BL	9	0.89	0.05	0.73	0.07	0.82	0.03	-0.001	0.973	1.52
Tyr	2nd der+MSC	7	0.87	0.12	0.76	0.16	0.87	0.10	-0.093	1.053	1.72
Leu	2nd der+SNV	10	0.91	0.11	0.78	0.17	0.83	0.12	0.050	0.910	1.51
Pro	2nd der+DT	10	0.88	0.06	0.72	0.09	0.80	0.06	0.019	0.741	1.54
Ile	1st der+MSC	8	0.84	0.05	0.75	0.06	0.88	0.03	0.017	0.942	2.00
Val	1st der+DT	8	0.88	0.06	0.81	0.07	0.78	0.06	0.007	0.278	1.58
Met	2nd der+BL	11	0.92	0.02	0.76	0.03	0.83	0.03	0.054	0.744	1.60
Glu	2nd der+BL	9	0.88	0.48	0.76	0.65	0.88	0.41	0.020	0.645	1.99
Ser	2nd der+DT	10	0.91	0.09	0.80	0.13	0.93	0.09	-0.005	1.150	2.08
Thr	1st der	8	0.82	0.05	0.71	0.06	0.83	0.04	0.010	0.835	1.50
His	2nd der+SNV	10	0.88	0.03	0.74	0.05	0.78	0.05	0.002	0.802	1.40

(Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (BL) Baseline correction; (DT) Detrending; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (MSC) Multiple scattering correction; (Phe) Phenylalanine; (Pro) Proline; (R) Pearson correlation coefficient; (RPD) Residual predictive deviation; (SEC) Standard error in calibration; (SECV) Standard error in cross validation; (SEP) Standard error in prediction; (Ser) Serine; (SNV) Standard normal variate; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine; (1st der) The first derivative; (2nd der) The second derivative.

REFERENCES

Agelet, L. E., Armstrong, P. R., Clariana, R. I., & Hurburgh, C. R. (2012). Measurement of single soybean seed attributes by near-infrared technologies. A comparative study. *Journal of Agricultural and Food Chemistry, 60*(34), 8314-8322. https://doi.org/10.1021/jf3012807.

Agelet, L. E., & Hurburgh, C. R., Jr. (2014). Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. *Talanta, 121*, 288-299. https://doi.org/10.1016/j.talanta.2013.12.038.

Berardo, N., Pisacane, V., Battilani, P., Scandolara, A., Pietri, A., & Marocco, A. (2005). Rapid detection of kernel rots and mycotoxins in maize by near-infrared reflectance spectroscopy. *Journal of Agricultural and Food Chemistry, 53*(21), 8128-8134. https://doi.org/10.1021/jf0512297.

Bishi, S. K., Kumar, L., Dagla, M. C., Mahatma, M. K., Rathnakumar, A. L., Lalwani, H. B., & Misra, J. B. (2013). Characterization of Spanish peanut germplasm (Arachis hypogaea L.) for sugar profiling and oil quality. *Industrial Crops and Products, 51*, 46-50. https://doi.org/10.1016/j.indcrop.2013.08.050.

Bishi, S. K., Lokesh, K., Mahatma, M. K., Khatediya, N., Chauhan, S. M., & Misra, J. B. (2015). Quality traits of Indian peanut cultivars and their utility as nutritional and functional food. *Food Chemistry, 167*, 107-114. https://doi.org/10.1016/j.foodchem.2014.06.076.

Caporaso, N., Whitworth, M. B., Fowler, M. S., & Fisk, I. D. (2018). Hyperspectral imaging for non-destructive prediction of fermentation index, polyphenol content and antioxidant activity in single cocoa beans. *Food Chemistry, 258*, 343-351. https://doi.org/10.1016/j.foodchem.2018.03.039.

Carbas, B., Machado, N., Oppolzer, D., Ferreira, L., Brites, C., Rosa, E. A. S., & Barros, A. (2020). Comparison of near-infrared (NIR) and mid-infrared (MIR) spectroscopy for the determination of nutritional and antinutritional parameters in common beans. *Food Chemistry, 306*, 125509. https://doi.org/10.1016/j.foodchem.2019.125509.

Chen, K., Li, C., & Tang, R. (2017). Estimation of the nitrogen concentration of rubber tree using fractional calculus augmented NIR spectra. *Industrial Crops and Products, 108*, 831-839. https://doi.org/10.1016/j.indcrop.2017.06.069.

Chopra, R., Folstad, N., Lyons, J., Ulmasov, T., Gallaher, C., Sullivan, L., McGovern, A., Mitacek, R., Frels, K., Altendorf, K., Killam, A., Ismail, B., Anderson, J. A., Wyse, D. L., & Marks, M. D. (2019). The adaptable use of Brassica NIRS calibration equations to identify pennycress variants to facilitate the rapid domestication of a new winter oilseed crop. *Industrial Crops and Products, 128*, 55-61. https://doi.org/10.1016/j.indcrop.2018.10.079.

Chu, X. (2011). *Molecular spectroscopy analytical technology combined with chemometrics and its applications*. Beijing: Chemical Industry Press.

Dettori, M., Arru, B., Azara, A., Piana, A., Mariotti, G., Camerada, M. V., Stefanelli, P., Rezza, G., & Castiglia, P. (2018). In the Digital Era, Is Community Outrage a Feasible Proxy Indicator of Emotional Epidemiology? The Case of Meningococcal Disease in Sardinia, Italy. *International Journal of Environmental Research and Public Health, 15*(7), 1512. https://doi.org/10.3390/ijerph15071512.

Dubinina, A., Lehnert, S., & Khomenko, O. (2014). Amino acid composition of protein and its biological value in seeds of peanut sorts widen in Ukraine. *Agriculture and Food, 2*(1), 501-510.

Faostat, F. (2020). Statistical databases. Retrieved from: http://www.fao.org/faostat/en/#data/QC Accessed.

Fearn, T. (2002). Assessing calibrations: sep, rpd, rer and r 2. *NIR news, 13*(6), 12-13.

Fernandez-Novales, J., Garde-Cerdan, T., Tardaguila, J., Gutierrez-Gamboa, G., Perez-Alvarez, E. P., & Diago, M. P. (2019). Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening. *Talanta, 199*, 244-253. https://doi.org/10.1016/j.talanta.2019.02.037.

Fox, G., Kelly, A., Sweeney, N., & Hocroft, D. (2011). Development of a single kernel NIR barley protein calibration and assessment of variation in protein on grain quality. *Journal of the Institute of Brewing, 117*(4), 582-586. https://doi.org/10.1002/j.2050-0416.2011.tb00507.x.

Fox, G. P., Wu, A., Yiran, L., & Force, L. (2013). Variation in caffeine concentration in single coffee beans. *Journal of Agricultural and Food Chemistry, 61*(45), 10772-10778. https://doi.org/10.1021/jf4011388.

Fraser, D. (2001). Near infra-red (NIR) light penetration into an apple. *Postharvest Biology and Technology, 22*, 191-194.

GB5009.5. (2016). *National Food Safety Standard - Determination of Protein in Foods*. Beijing: Standard Press of China.

GB5009.6. (2016). *National Food Safety Standard - Determination of Fat in Foods*. Beijing: Standard Press of China.

GB5009.8. (2016). *National food safety standard - Determination of fructose, glucose, sucrose, maltose and lactose in foods* Beijing: Standard Press of China.

GB5009.124. (2016). *National Food Safety Standard - Determination of Amino Acid in Foods*. Beijing: Standard Press of China.

Gong, A., Shi, A., Liu, H., Yu, H., Liu, L., Lin, W., & Wang, Q. (2018). Relationship of chemical properties of different peanut varieties to peanut butter storage stability. *Journal of Integrative Agriculture, 17*(5), 1003-1010. https://doi.org/10.1016/s2095-3119(18)61919-7.

Hom, N. H., Becker, H. C., & Möllers, C. (2006). Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. *Euphytica, 153*(1-2), 27-34. https://doi.org/10.1007/s10681-006-9195-3.

Hourant, P., Baeten, V., Morales, M. T., Meurens, M., & Aparicio, R. (2000). Oil and fat classification by selected bands of near-infrared spectroscopy. *Applied Spectroscopy, 54*(8), 1168-1174. https://doi.org/10.1366/0003702001950733.

Jiang, H., Duan, N., & Ren, X. (2006). *Specification of peanut germplasm resources*. Beijing: China Argricultrue Press.

Jiang, H. Y., Zhu, Y. J., Wei, L. M., Dai, J. R., Song, T. M., Yan, Y. L., & Chen, S. J. (2007). Analysis of protein, starch and oil content of single intact kernels by near infrared reflectance spectroscopy (NIRS) in maize (Zea mays L.). *Plant Breeding, 126*(5), 492-497. https://doi.org/10.1111/j.1439-0523.2007.01338.x.

Klevorn, C. M., Dean, L. L., & Johanningsmeier, S. D. (2019). Metabolite Profiles of Raw Peanut Seeds Reveal Differences between Market-Types. *Journal of Food Science, 84*(3), 397-405. https://doi.org/10.1111/1750-3841.14450.

Kutsanedzie, F. Y. H., Hao, L., Yan, S., Ouyang, Q., & Chen, Q. (2018). Near infrared chemo-responsive dye intermediaries spectra-based in-situ quantification of volatile organic compounds. *Sensors and Actuators B: Chemical, 254*, 597-602. https://doi.org/10.1016/j.snb.2017.07.134.

Leng, T., Li, F., Xiong, L., Xiong, Q., Zhu, M., & Chen, Y. (2020). Quantitative detection of binary and ternary adulteration of minced beef meat with pork and duck meat by NIR combined with chemometrics. *Food Control, 113*, 107203. https://doi.org/10.1016/j.foodcont.2020.107203.

Li, Y., Zhang, R., Qin, X., Liao, Y., Siddique, K. H. M., & Prasad, D. M. (2018). Changes in the protein and fat contents of peanut (Arachis hypogaea L.) cultivars released in China in the last 60 years. *Plant Breeding, 137*(5), 746-756. https://doi.org/10.1111/pbr.12621.

Lima, A. B. S. d., Batista, A. S., Jesus, J. C. d., Silva, J. d. J., Araújo, A. C. M. d., & Santos, L. S. (2020). Fast quantitative detection of black pepper and cumin adulterations by near-infrared spectroscopy and multivariate modeling. *Food Control, 107*, 106802. https://doi.org/10.1016/j.foodcont.2019.106802.

Liu, N., Parra, H. A., Pustjens, A., Hettinga, K., Mongondry, P., & van Ruth, S. M. (2018). Evaluation of portable near-infrared spectroscopy for organic milk authentication. *Talanta, 184*, 128-135. https://doi.org/10.1016/j.talanta.2018.02.097.

Natsuga, M., Sue, Y., Ikeda, T., Egashira, H., Akazawa, T., & Ura, N. (2007). Development and evaluation of dedicated bench-type near infrared spectrometers to estimate the flavour of green soybeans from single pod measurements. *Journal of Near Infrared Spectroscopy, 15*(5), 327-332. https://doi.org/10.1255/jnirs.741.

Nicolaï, B. M., Beullens, K., Bobelyn, E., Peirs, A., Saeys, W., Theron, K. I., & Lammertyn, J. (2007). Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review. *Postharvest Biology and Technology, 46*(2), 99-118. https://doi.org/10.1016/j.postharvbio.2007.06.024.

Pérez-Rodríguez, F., Skandamis, P., & Valdramidis, V. (2018). *Quantitative Methods for Food Safety and Quality in the Vegetable Industry*. New York: Springer.

Perry, S. E., & Wang, H. (2003). Rapid isolation of Arabidopsis thaliana developing embryos. *Biotechniques, 35*(2), 278-282. https://doi.org/10.2144/03352bm06.

Quintelas, C., Mesquita, D. P., Ferreira, E. C., & Amaral, A. L. (2019). Quantification of pharmaceutical compounds in wastewater samples by near infrared spectroscopy (NIR). *Talanta, 194*, 507-513. https://doi.org/10.1016/j.talanta.2018.10.076.

Radhakrishnan, R., Pae, S.-B., Kang, S.-M., Lee, B.-K., Lee, I.-J., & Baek, I.-Y. (2014). An evaluation of amino acid, fatty acid and isoflavone composition in Korean peanut (Arachis hypogaea l.) seeds to improve the nutritional quality of breeding lines. *J. Korean Soc. Appl. Bi., 57*(3), 301-305. https://doi.org/https://doi.org/10.1007/s13765-014-4044-x.

Sarvamangala, C., Gowda, M. V. C., & Varshney, R. K. (2011). Identification of quantitative trait loci for protein content, oil content and oil quality for groundnut (Arachis hypogaea L.). *Field Crops Research, 122*(1), 49-59. https://doi.org/10.1016/j.fcr.2011.02.010.

Shao, Y., Cen, Y., He, Y., & Liu, F. (2011). Infrared spectroscopy and chemometrics for the starch and protein prediction in irradiated rice. *Food Chemistry, 126*(4), 1856-1861. https://doi.org/10.1016/j.foodchem.2010.11.166.

Shen, F., Niu, X., Yang, D., Ying, Y., Li, B., Zhu, G., & Wu, J. (2010). Determination of amino acids in Chinese rice wine by fourier transform near-infrared spectroscopy. *Journal of Agricultural and Food Chemistry, 58*(17), 9809-9816. https://doi.org/10.1021/jf1017912.

Shin, E.-C., Craft, B. D., Pegg, R. B., Phillips, R. D., & Eitenmiller, R. R. (2010). Chemometric approach to fatty acid profiles in Runner-type peanut cultivars by principal component analysis (PCA). *Food Chemistry, 119*(3), 1262-1270. https://doi.org/10.1016/j.foodchem.2009.07.058.

Sundaram, J., Kandala, C. V., Holser, R. A., Butts, C. L., & Windham, W. R. (2010). Determination of in-shell peanut oil and fatty acid composition using near-infrared reflectance spectroscopy. *Journal of the American Oil Chemists' Society, 87*(10), 1103-1114. https://doi.org/10.1007/s11746-010-1589-7.

Tallada, J. G., Palacios-Rojas, N., & Armstrong, P. R. (2009). Prediction of maize seed attributes using a rapid single kernel near infrared instrument. *Journal of Cereal Science, 50*(3), 381-387. https://doi.org/10.1016/j.jcs.2009.08.003.

Tillman, B. L., Gorbet, D. W., & Person, G. (2006). Predicting Oleic and Linoleic Acid Content of Single Peanut Seeds using Near-Infrared Reflectance Spectroscopy. *Crop Science, 46*(5), 2121. https://doi.org/10.2135/cropsci2006.01.0031.

Vallone, M., Alleri, M., Bono, F., & Catania, P. (2019). Quality evaluation of grapes for mechanical harvest using vis NIR spectroscopy. *Agricultural Engineering International: CIGR Journal, 21*(1), 140-149.

Wang, L., Wang, Q., Liu, H., Liu, L., & Du, Y. (2013). Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy. *Journal of Food Science and Technology, 93*(1), 118-124. https://doi.org/10.1002/jsfa.5738.

Wang, Q. (2016). *Peanuts: Processing technology and product development*. Cambridge: Academic Press.

Wang, Q. (2018). *Peanut processing characteristics and quality evaluation*. New York: Springer.

Wang, Q., Liu, H., Shi, A., Hu, H., Liu, L., Wang, L., & Yu, H. (2017). Review on the processing characteristics of cereals and oilseeds and their processing suitability evaluation technology. *Journal of Integrative Agriculture, 16*(12), 2886-2897. https://doi.org/10.1016/s2095-3119(17)61799-4.

Williams, P., & Norris, K. (1987). *Near-infrared technology in the agricultural and food industries*. Eagan: American Association of Cereal Chemists, Inc.

Workman Jr, J., & Weyer, L. (2012). *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*. Boca Raton: CRC press.

Xu, S., Wang, J. J., Wei, Y., Deng, W. W., Wan, X., Bao, G. H., Xie, Z., Ling, T. J., & Ning, J. (2019). Metabolomics based on uhplc-orbitrap-ms and global natural product social molecular networking reveals effects of time scale and environment of storage on the metabolites and taste quality of raw pu-erh tea. *Journal of Agricultural and Food Chemistry, 67*(43), 12084-12093. https://doi.org/10.1021/acs.jafc.9b05314.

Xu., Fan, S., Liu, J., Liu, B., Tao, L., Wu, J., Hu, S., Zhao, L., Wang, Q., & Wu, Y. (2019). A calibration transfer optimized single kernel near-infrared spectroscopic method. *Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy, 220*, 117098. https://doi.org/10.1016/j.saa.2019.05.003.

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. https://doi.org/10.1002/ejlt.201900031.

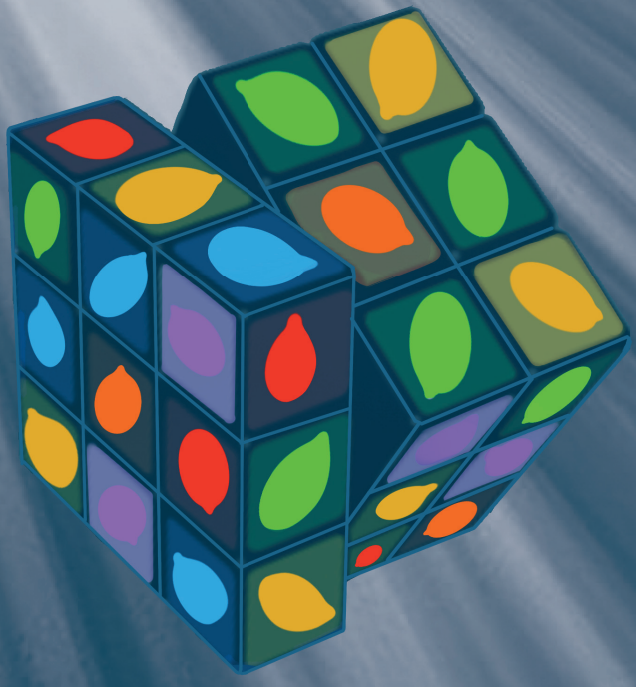
Ye, M., Gao, Z., Li, Z., Yuan, Y., & Yue, T. (2016). Rapid detection of volatile compounds in apple wines using FT-NIR spectroscopy. *Food Chemistry, 190*, 701-708. https://doi.org/10.1016/j.foodchem.2015.05.112.

Yu, H., Liu, H., Wang, N., Yang, Y., Shi, A., Liu, L., Hu, H., Mzimbiri, R. I., & Wang, Q. (2016). Rapid and visual measurement of fat content in peanuts by using the hyperspectral imaging technique with chemometrics. *Analytical Methods, 8*(41), 7482-7492. https://doi.org/10.1039/c6ay02029a.

Yu, H., Liu, H., Wang, Q., & van Ruth, S. (2020). Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. *LWT - Food Science and Technology*, 109398. https://doi.org/10.1016/j.lwt.2020.109398.

Yu, H., Wang, Q., Shi, A. M., Yang, Y., Liu, L., Hu, H., & Liu, H. Z. (2017). Visualization of protein in peanut using hyperspectral image with chemometrics. *Spectroscopy and Spectral Analysis, 37*(3), 853-858. https://doi.org/10.3964/j.issn.1000-0593(2017)03-0853-06.

Zaaboul, F., Raza, H., Chen, C., & Liu, Y. (2018). Characterization of peanut oil bodies integral proteins, lipids, and their associated phytochemicals. *Journal of Food Science, 83*(1), 93-100. https://doi.org/10.1111/1750-3841.13995.



CHAPTER 4

An explorative study on the relationships between the quality traits of peanut varieties and their peanut butters

This chapter has been published as:
Yu, H., Liu, H., Erasmus, S. W., Zhao S., Wang, Q., & van Ruth, S. M. (2021). An explorative study on the relationships between the quality traits of peanut varieties and their peanut butters. LWT - Food Science and Technology, 151, 112068.

ABSTRACT

Peanut varieties have their own distinct characteristics, which also affect the properties of the processed-products. Knowledge on these effects can assist peanut processors to select certain varieties for specific products. Therefore, the multivariate relationships between the quality traits of peanut kernels and their peanut butters were explored in this study. Peanut butters were manufactured from forty peanut varieties with detailed component analysis. The volatile compounds, colour, texture, rheology, and particle size distributions of peanut butters and their relationships with peanut varieties were comprehensively analysed. The results showed that peanut butters prepared from varieties with lower fat, higher protein, and higher sucrose contents have higher firmness, yield stress, and *G*" values. It was also suggested that amino acids likely play a major role in the formation of the distinct peanut butter aroma and colour. This research study provided a detailed analysis approach that can be used by processing enterprises to select the most suitable varieties to produce peanut butters that address different commercial needs.

Keywords: Cluster analysis; Correlation analysis; Principal component analysis; Texture and rheology; Volatile compounds

4.1. INTRODUCTION

Peanut butters are dense suspensions of solid flecks diffused in a continuous oil phase (Norazatul Hanim et al., 2016). Globally, it is one of the most consumed and important plant-based spreads which has a special flavour and mouthfeel (Gong et al., 2018), while it also offers health benefits such as inhibiting the risk of gastric cancer (Hashemian et al., 2017). Nowadays, the industrial scale production of peanut butter is extensively developed and generally processed by subjecting raw peanuts to roasting and grinding. Peanuts form the material basis of peanut butter and hence the quality of the intact peanuts can subsequently affect the quality characteristics of its resulting peanut butter.

Peanuts are identified as one of the leading produced, processed, and traded nuts in the world (Wang, 2018). There are more than 8,000 peanut varieties and strains worldwide which differ considerably in their composition (Wang, Liu, Shi, et al., 2017). The quality requirement of peanuts to produce different types of products like peanut tofu, peanut oil, and peanut butter vary substantially. It is well-known that the amino acids contents such as cystine and lysine play a vital role in peanut protein (Wang, Liu, Liu, et al., 2017) and peanut tofu (Chen et al., 2020). It has also been discovered that peanuts with higher oleic acid contents result in better peanut oils after processing (Wang, 2018). The acceptability of peanut butter is linked to various quality traits that includes its volatile compounds, colour, texture, rheology, and particle size distributions (Shakerardekani et al., 2013; Starowicz & Zieliński, 2019). Most studies have only focused on one or two quality traits of peanut butters from a few varieties. For example, Mohd Rozalli et al. (2015) focused on the diameter of particles and the dynamic rheological qualities of peanut butters, showing that peanut butters have elastic properties. Norazatul Hanim et al. (2016) discovered that two peanut varieties resulted in different particle size distributions of peanut butters. The volatiles of two different peanut butters have also been studied, indicating that pyrazines are the main aroma components of peanut butters (Lou et al., 2009). Some scientists (Dhamsaniya et al., 2012; Gong et al., 2018) have already conducted research on the impacts of different fat, protein, and fatty acids contents of peanuts on the sensory properties of peanut butters and its stability during the storage. Although some relationships between peanuts and the resulting peanut butters have been revealed, the picture is not yet complete.

More in-depth knowledge about the relationships and interactions that exist between quality traits of peanut kernels and peanut butters is valuable for the development of new, nutritious, and tasteful products. This quality appraisal of peanuts is not only necessary to select product-specific varieties for satisfying the commercial demands of markets, but also to better allocate how the raw material should be used to increase the sustainability of the industry. Hence, the first aim of the research was to comprehensively analyse

the properties of forty peanut varieties, i.e. physical traits and composition, as well as the properties and quality traits of the resulting peanut butters. The latter included measurements of the volatile compounds, colour, texture, rheology, and particle size distributions. Subsequently, univariate comparisons and multivariate comparisons were conducted on peanut kernels and peanut butters to explore the relationships between traits for each group. Lastly, the relationships between peanut varieties and peanut butters were also elucidated.

4.2. MATERIALS AND METHODS

4.2.1. Peanuts and peanut butters

Forty peanut varieties were collected from different provinces in China (the details are listed in **Figure S4.1**). After collection, the unshelled peanuts (approximately 5 kg per sample) were kept in refrigerated storage units (4°C, Yuandong Co., Ltd., Tianjing, China) until further processing. Peanut butter was manufactured as follows. The peanuts per variety were shelled and about 500 g mature plump peanut kernels (displaying no mildew or damage) were selected. The peanut kernels were then spread in a single layer on a baking sheet and roasted for 30 min at 160°C in a commercial oven with a supper and lower heating function (ROBAM Co., Ltd, Hangzhou, China). After roasting, the peanut kernels were removed from the oven and left to cool at room temperature (21°C) before they were peeled for grinding. Peanut butter was produced by grinding the roasted peanut kernels twice in a colloid mill (Langfang Tongyong Machinery Co., Ltd., Hebei, China). The first grinding, that lasted for 10 minutes, was to make sure that all peanut kernels were ground. The second grinding, that lasted for 30 seconds, was performed to ensure a final smooth product. Ultimately, 400 g peanut butter was produced for each peanut variety. All peanut butters were stored at room temperature (21°C) until analysis. The whole procedure and analyses performed in this study are presented in **Figure 4.1**.

4.2.2. Physical analysis of the peanuts

The dimensions (width and length) of the peanuts were determined using a Vernier Caliper (Shanghai Tool Factory Co., Ltd., Shanghai, China) (Yu, Liu, Erasmus, et al., 2020). The width (mm) and the length (mm) of each peanut kernel are respectively the crosswise and the longitudinal distance of the mature seed (Jiang et al., 2006). Ten mature plump peanut kernels of each variety, without mildew or damage, were selected for width and length measurements. The total weight of one hundred mature plump kernels per variety were considered as weight (g). All varieties were analysed in duplicate and the average values were calculated for all parameters.

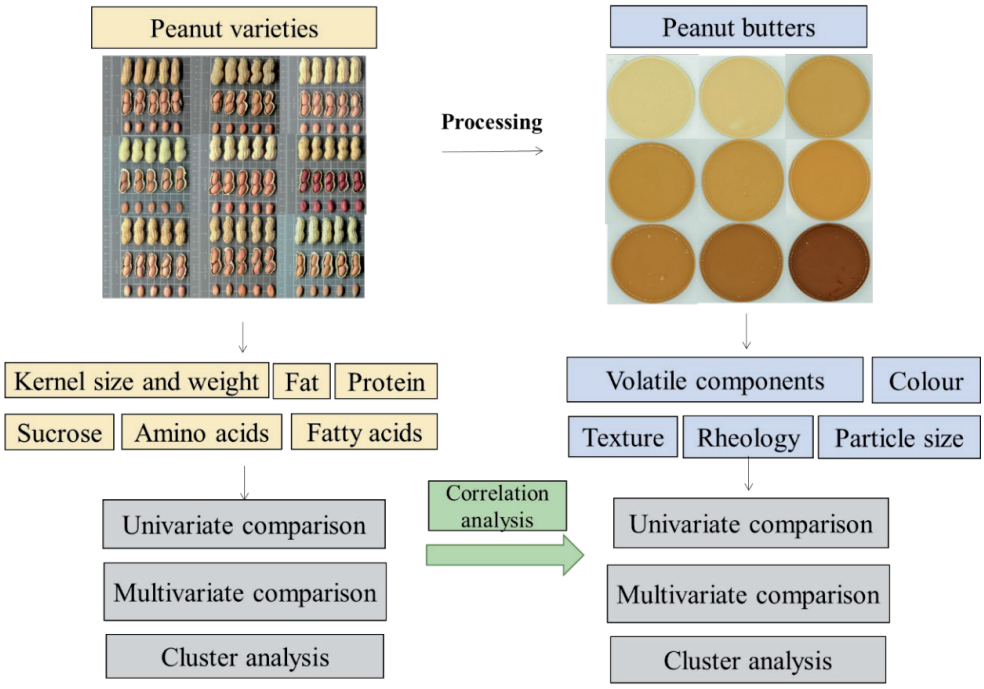


Figure 4.1. The technology roadmap for the sourcing and processing of the peanuts and subsequent analyses conducted.

4.2.3. Compositional analysis of the peanuts

The fat contents of the peanuts were determined by the Soxhlet extraction procedure (GB5009.6, 2016). The protein contents were established using the Kjeldahl procedure (GB5009.5, 2016) with a conversion coefficient of 5.46. Amino acids in the peanuts were evaluated following the standard method (GB5009.124, 2016), while sucrose contents were examined according to the standard method (GB5009.8, 2016) and fatty acid compositions were determined by the corresponding standard method (GB5009.168, 2016). Each analysis was performed in duplicate for each peanut variety with the average values calculated for all parameters.

4.2.4. Analysis of the volatile compounds of the peanut butters

Headspace solid-phase micro-extraction (HS-SPME) was utilised to determine the volatile compounds of the peanut butters. A SPME fibre (50/30 µm divinylbenzene/Carboxen/polydimethylsiloxane, Stableflex, Supelco Co., Bellefonte, the USA) was pre-treated at 270°C for 30 min prior to the first measurement. Each sample (5 g) was loaded into a 20 mL glass vial sealed with an aluminium cover and a Teflon septum. For quantification,

50 µL 1,2,3-trichloropropane (0.5 mg/mL in methanol, Sinopharm Chemical Reagent Co., Ltd., Beijing, China) was used as the internal standard. Each sample was pre-equilibrated for 10 min at 55°C in a shaking incubator. The SPME holder with the fibre was then embedded into the sample vial and exposed to the headspace for 40 min to absorb the volatile compounds. These compounds were then desorbed in the hot injection part of gas chromatography–mass spectrometry (GC-MS) for 150 s at 260°C. The GC system (Angilent 7890B, Agilent Technologies, Santa Clara, CA) and a mass selective analyser (Angilent 5977B) equipped with a VF-MAX column (30 m × 0.25 mm ID, 0.25 µm film thickness; Angilent CP9205, Agilent Technologies, Santa Clara, CA) were used for the GC-MS analysis. The analysis was conducted in splitless mode, using helium as the carrier gas at a flow rate of 1 mL/min. The setting temperature of the analyser was 250°C, and the oven temperature program was set to start at 40°C for 5 min and programmed to increase with 5 °C/min to 250°C for holding 5 min. The electron impact ionization mode (70 eV) was used to record the mass spectrum and scan the mass range from 35 to 500 m/z. The ion source temperature was held at 230°C. The volatile compounds were tentatively identified by comparing the sample results with the mass spectral database from the National Institute of Standards and Technology (NIST) library. In addition, an n-alkane mixed standard (C7-C40) (0.5 mg/mL in N-hexane, o2si Smart Solutions, North Charleston, American) was run in the same conditions to obtain the retention indices (RI). The RI were calculated as stated in the formula (Eq. (1)).

$$RI_x = 100n + 100 \times {(tR_x - tR_n)}\Big/{(tR_{n+1} - tR_n)} \quad (1)$$

where the retention time (tR) of $tR_n < tR_x < tR_{n+1}$; n = number of atom carbon; x = one of volatile aroma compounds. Each sample was analysed in duplicate and the average values were calculated for all parameters.

4.2.5. Assessment of the colour of the peanut butters

The colour properties of the peanut butters were obtained by a CS-600 portable colour spectrophotometer (CHNSpec Technology Co., Ltd, Hangzhou, China). About 2 g peanut butters were placed into a round quartz cell (30 mm diameter × 40 mm height) for assessment. Three different properties representing different colours were detected: L*, darkness - lightness (0 - 100); a*, greenness - redness (-128 - 127); b*, blueness - yellowness (-128 - 127). The procedure was repeated in triplicate for each peanut butter and the average values were calculated for all parameters.

4.2.6. Assessment of the texture of the peanut butters

The texture of the peanut butters was assessed using a Texture Analyser (TA-XTplus, Micro Stable System Co., UK) to apply 30 mm penetration on the peanut butters at a speed of 1 mm/s and a trigger force of 5 g. The back-extrusion apparatus was equipped with a locating base plate, a sample container (50 mm internal diameter), and a compression disc (35 mm diameter). To preciously measure the textural qualities of the peanut butters, samples were placed into the container at the same volume level (75% full). The exerted force (up and down) was automatically measured. The texture of the peanut butters was explained by the following parameters: firmness (g), cohesiveness (g), consistency (g·s), and index of viscosity (g·s). Firmness is defined as the force at the maximum distance of a penetration cycle, while cohesiveness is the maximum negative force (expressed as positive values). Consistency and the index of viscosity are the area of positive and negative zone (expressed as positive values), respectively. Each sample was analysed in duplicate and the average values were calculated for all parameters.

4.2.7. Assessment of the rheology of the peanut butters

The rheological properties of peanut butters were characterised by a controlled stress rheometer (HR-2, TA Instruments, New Castle, the USA) armed with a crosshatched plate (40 mm diameter, plate geometry with 1 mm gap). About 2 g peanut butter was loaded onto the plate for analysis. The rheological properties were detected through two aspects: shear rate sweep and dynamic oscillatory tests. For the shear rate sweep test, the steady state detections were carried out at a shear rate range of 1 - 300/s after 2 min of conditioning. The flow behaviour was modelled using the Herschel–Bulkley's model (Eq. (2)) (Taghizadeh & Razavi, 2009). It is generally used to indicate the rheological properties of foods to provide rheological constants of food products.

$$\sigma = \sigma_0 + K \times r^n \tag{2}$$

Where δ stands for shear stress (Pa), δ_0 stands for yield stress (Pa), K stands for consistency coefficient (Pa·S^n), r stands for shear rate (S^{-1}), and n stands for flow behaviour index (Zhou et al., 2017).

The dynamic oscillatory experiment was conducted to collect dynamic rheological behaviours including the storage modulus (G') and the loss modulus (G'') through frequency sweep. A constant angular frequency (1 Hz) was used to identify the linear viscoelastic range (LVR) of peanut butter. The frequency was increased from 0.1 to 100 Hz, and all tests were conducted within the LVR. The G' and G'' were modelled as a power function of oscillatory frequency (ω) (Eq. (3), (4)), as generally used for indicating the viscoelastic behaviour of peanut butter, which was used to indicate firmness and viscosity.

$$G' = a \times \omega^b \quad (3)$$

$$G'' = c \times \omega^d \quad (4)$$

The power law coefficients, *a/c* (kPa $Hz^{-b/d}$), illustrate the amount of G' and G" respectively at a given frequency. The power law exponents, *b/d* (×100), illustrate the slope of the relationships between the modulus and frequency (Liu et al., 2019; Resch & Daubert, 2002). Each peanut butter sample was analysed in duplicate and the average values were calculated for all parameters.

4.2.8. Analysis of the particle size distributions of the peanut butters

The particle size distributions of the peanut butters were detected by a Malvern 3000 Mastersizer (Malvern Instruments Ltd, Worcerstershire, U.K.) equipped with a sample input cell (Hydro 2000 MU). Each sample (0.1 g) was weighed into a 25 mL test tube and 15 mL of acetone (Sinopharm Chemical Reagent Co., Ltd, Beijing, China) as the diluter was appended to disperse the sample by a vortex mixer (Scientific Industries, Inc, Bohemia, the USA). After that, the diluted sample was placed into the input cell stuffed with 600 mL deionized water until the obscuration range was within 0.10 to 0.15. The obscuration indicates the magnitude of light obscured by the sample as a result of the scattering and absorbing effects. The refractive indexes of acetone and the peanut butter were 1.376 and 1.500, respectively (Norazatul Hanim et al., 2016). The acquired data were analysed using the Malvern software (Malvern Instruments Ltd, Worcerstershire, U.K.) according to the Mie Scattering theory. Each sample was analysed three times and the average values were calculated for all diameter parameters. The polydispersity value indicating the particle size was specified in Eq. (5).

$$Polydispersity = {(D90 - D10)}/{D50} \quad (5)$$

Where D10, D50, and D90 are the average value of diameter which falls below 10th, 50th, and 90th percentile of particle size distributions.

4.2.9. Data analysis

Univariate comparisons were carried out on the data of the peanut kernels and peanut butters to identify statistical features, including average, standard deviation (SD), minimum (min), maximum (max), first quartile (Q1), and third quartile (Q3) values. Principal component analysis (PCA) was used to explore the relationships of average values of quality traits of peanut varieties and the key characteristics of peanut butters. Unsupervised clustering was explored using K-means procedures. The distance between each sample and the nearest mean was determined to obtain the different groups. Kruskal-Wallis tests

were applied to assess if there were any significant differences between different peanut butter groups because the data were not normally distributed as shown by the Shapiro-Wilk tests. Pairwise comparisons were implemented by the Mann-Whitney U tests to evaluate differences in various traits of peanut butters between different groups. Correlation analysis was used to evaluate the strength of the relationships between univariate qualities of the peanuts and univariate results and PCA outcomes of the peanut butters. R 4.0.1 statistical software (R Foundation for Statistical Computing, Austria) was used for all calculations.

4.3. RESULTS AND DISCUSSION

4.3.1. The traits of the peanut varieties: univariate comparisons

The quality traits of forty peanut varieties are shown in **Table 4.1**. The dimensions of the peanut kernels varied between 14.4 - 20.5 mm (length) and 8.5 - 11.6 mm (width). The weight of the peanut kernels was mainly determined by seed size (as represented by its dimensions), which meant that smaller seed size varieties had a lower weight, and vice versa (Zhang et al., 2019). For example, the weight of SH8 (length 14.4 ± 1.3 mm and width 8.5 ± 0.5 mm) was 43.3 ± 1.2 g, while the weight of WH8 (length 19.6 ± 1.0 mm and width 11.3 ± 0.5 mm) was 101.7 ± 2.3 g. The range of fat and protein in the peanut kernels were 42.6 - 52.9 g/100 g and 19.3 - 26.4 g/100 g, respectively, which was similar to previously reported results (Gong et al., 2018; Nayak et al., 2020; Wang, 2016). The content of sucrose ranged from 3.3 - 7.9 g/100 g, which was consistent with the previous studies (Bishi et al., 2013; Wang, 2018). Among all varieties, the sucrose content in JHT1 was up to 7.9 ± 0.2 g/100 g. Among the measured amino acids, Arginine (Arg), Aspartic acid (Asp), and Glutamic acid (Glu) were the three most abundant compounds and similar results have been reported in previous research (Klevorn et al., 2019; Wang et al., 2013).

In this study, twelve high oleic acid peanut varieties (DF06, HY661, HY917, HY962, HY963, JH16, JH18, JH19, WH25, YH37, YH65, and YH76) were used in addition to regular varieties since they, as raw materials for manufacture can extend the shelf life of final products (Gong et al., 2018) and boost a product's nutritional quality (Zhao, Shi, et al., 2019). Therefore, a considerable variation in fatty acids composition (oleic acid, 14.82 - 43.83 g/100 g; linoleic acid 1.24 - 23.61 g/100 g) was noted. The wide variation of the quality traits suggested an abundant diversity among the forty peanut varieties.

Table 4.1. The univariate descriptive analysis of the quality traits of peanut kernel varieties (n = 40).

Traits	Average	Min	Max	SD	Q1	Q3
Length (mm)	17.5	14.4	20.5	1.7	16.0	18.7
Width (mm)	10.2	8.5	11.6	0.8	9.7	10.7
Weight (100 kernels, g)	73.6	43.4	101.7	14.7	61.3	86.1
Fat (g/100 g)	48.3	42.	52.9	3.0	46.1	50.7
Protein (g/100 g)	23.6	19.3	26.4	1.6	22.6	24.9
Sucrose (g/100 g)	4.8	3.3	7.9	1.0	4.1	5.5
Glu (g/100 g)	5.03	3.42	6.01	0.53	4.79	5.39
Asp (g/100 g)	3.08	2.27	3.84	0.35	2.88	3.27
Arg (g/100 g)	2.99	1.89	3.56	0.35	2.76	3.25
Leu (g/100 g)	1.77	1.35	2.12	0.20	1.63	1.90
Gly (g/100 g)	1.54	1.09	1.99	0.19	1.47	1.64
Ser (g/100 g)	1.28	0.89	1.94	0.21	1.16	1.39
Thr (g/100 g)	0.64	0.38	0.79	0.08	0.58	0.70
Cys (g/100 g)	0.37	0.28	0.60	0.08	0.33	0.41
Phe (g/100 g)	1.29	0.84	1.83	0.17	1.21	1.37
Ala (g/100 g)	0.98	0.82	1.11	0.08	0.93	1.05
Met (g/100 g)	0.32	0.19	0.46	0.07	0.28	0.37
Val (g/100 g)	1.00	0.63	1.34	0.17	0.87	1.11
Pro (g/100 g)	1.01	0.71	1.53	0.17	0.88	1.07
Tyr (g/100 g)	1.00	0.63	1.25	0.16	0.87	1.11
Lys (g/100 g)	0.98	0.77	1.20	0.09	0.92	1.03
His (g/100 g)	0.55	0.40	0.66	0.06	0.51	0.60
Ile (g/100 g)	0.84	0.58	1.31	0.16	0.75	0.90
Oleic acid (g/100 g)	25.56	14.82	43.83	9.68	17.86	36.48
Linoleic acid (g/100 g)	12.91	1.24	23.61	7.61	2.18	18.46
Palmitic acid (g/100 g)	4.75	2.20	7.02	1.48	2.98	6.01

(Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; Cys (Cysteine); (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Max) maximum; (Met) Methionine; (Min) Minimum; (O/L) Oleic acid/ Linoleic acid; (Phe) Phenylalanine; (Pro) Proline; (Q1) Quartile 1; (Q3) Quartile 3; (Ser) Serine; (SD) Standard deviation; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

4.3.2. The traits of the peanut varieties: multivariate comparisons

The PCA biplot (**Figure 4.2**) shows that the samples were clustered in three groups based on the quality traits of peanut varieties in **Table 4.1**. The sum of PC1 and PC2 was 48%, accounting for half of the total variation. The "contrib" values represent the contributions of these traits on explaining the variations retained by the first two PCs. Among them, amino acids such as Alanine (Ala), Glutamic acid (Glu), and Lysine (Lys) and fatty acids such as oleic acid had the greatest contributions to PC1 and PC2, respectively. According to the K-means results, all high oleic acid peanut varieties (cluster 2) distributed in the negative half along PC2, while the regular peanut varieties (cluster 1 and 3) were located in the opposite direction because oleic acid had a negative correlation with

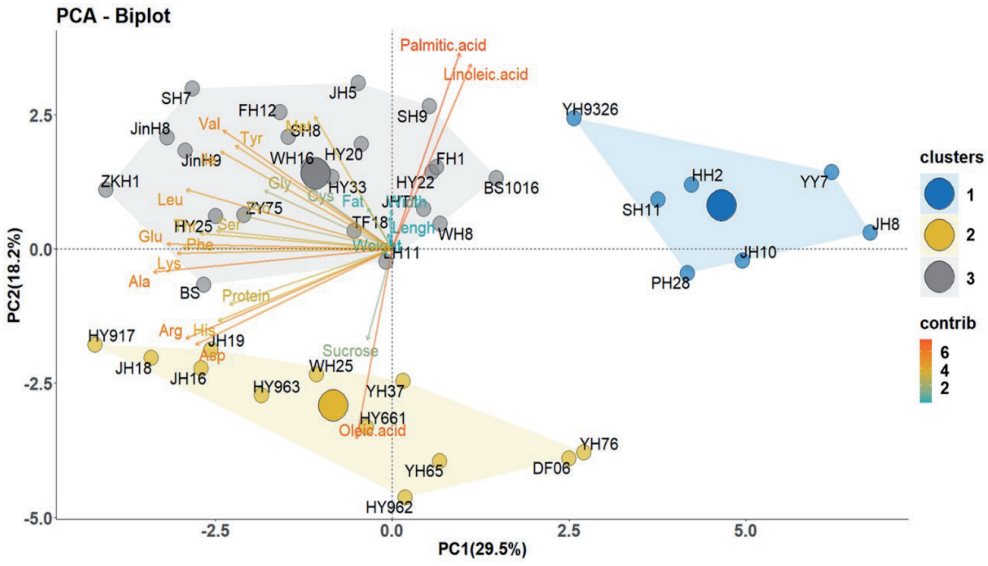


Figure 4.2. The principal component analysis (PCA) biplot of different peanut varieties (n=40). The different colored polygons indicate the different sample clusters and the arrows indicate the contributions of the quality traits. (Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (Cys) Cysteine; (contrib) Contribution; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (Phe) Phenylalanine; (Pro) Proline; (Ser) Serine; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

palmitic acid and linoleic acid in peanut samples (Yu, Liu, Wang, et al., 2020). The regular peanut varieties (cluster 1 and 3) were separated according to the contributions of amino acids and protein content along PC1. Cluster 3 had closer association with amino acids than cluster 1, especially for Ala, Arginine (Arg), Glu, Leucine (Leu), and Lys. In contrast to above variables, the fat and physical characteristics of the peanut varieties were not the main contributions for PC1 and PC2. This was mainly because there were no remarkable differences in fat content and physical characteristics between the varieties.

4.3.3. The traits of the peanut butters: univariate comparisons

All parameters of the studied peanut butters are also shown in **Table 4.2**. The unique aroma is the key quality of peanut butter. The attractive aroma of peanut butter is formed through reactions such as the Maillard reaction during the peanut roasting and grinding processes (Starowicz et al., 2019; Wang, Adhikari, et al., 2017). About 41 volatile compounds, including pyrazines, furans, esters, pyridines, and aldehydes, were detected for each peanut butter (Appendix supplement can be found online at https://doi. org/10.1016/j.lwt.2021.112068, Appendix supplement 2). As shown in **Figure 4.3a**, pyrazines and furans had obvious quantitative advantages than other compounds in the

peanut butters, which was similar to the previous research (Lou et al., 2009). The link lines between the volatile compounds and the peanut butters indicate that the contents of volatile compounds exceeded 5 mg/kg in **Figure 4.3a**. The thicker the line, the higher the contents. The volatile compounds varied greatly between different varieties. Among them, HY20, HY25, JH16, JH18, and LH11 presented higher concentrations of volatile compounds than other varieties.

It is likely that the intensity of volatile compounds was not only related to the concentrations of compounds, but also related to the odour threshold (OT) (Tamura et al., 2010). Therefore, both the concentration and OT should be considered to identify the compounds that commonly make the most contributions to the overall aroma of the peanut butters. Normally, the compounds are regarded as the effective compounds when the ratio of concentration/OT value is over one. Among them, pyrazines, usually with nutty and roast aromas, were the main volatile compounds of peanut butter, accounting for about 30% of the total volatile concentrations. Many studies have reported that pyrazines are the primary volatiles of peanut products (Baker et al., 2003; Li & Hou, 2018). In line with this, 2,5-dimethylpyrazine (RI = 915), 2,3,5-trimethylpyrazine (RI = 999), and 3-ethyl-2,5-dimethylpyrazine (RI = 1041) were the top three pyrazines detected in this study. As shown in **Table 4.2**, the average value of 2,5-dimethylpyrazine in peanut butters was 5.10 ± 2.03 mg/kg with the OT in vegetable oil 2.6 mg/kg (van Gemert, 2011); the ratio of concentration/OT value ≈ 2. The average value of 2,3,5-trimethylpyrazine was 2.17 ± 0.99 mg/kg and the OT in vegetable oil was 0.29 mg/kg (van Gemert, 2011); ratio of concentration/OT value ≈ 7, while the average value of 3-ethyl-2,5-dimethylpyrazine (RI = 1041) was 1.41 ± 0.66 mg/kg and the OT in oil was 0.024 mg/kg (van Gemert, 2011); the ratio of concentration/OT value ≈ 59. 2,5-Dimethylpyrazine and 2,3,5-trimethylpyrazine not only have a strong correlation with roasted peanut aroma (Baker et al., 2003; Wang, Adhikari, et al., 2017), but 2,5-Dimethylpyrazine can also be used as the best predictor of roasted aroma.

Apart from the above-mentioned pyrazines as effective compounds, 2-acetylpyrrole (RI = 1564) and furaneol (RI = 1629), also known as typical Maillard reaction products, have been stated to augment the roasty aroma of peanut oil and nut products (Arsa & Theerakulkait, 2018; Liu et al., 2011; van Gemert, 2011). The average value of 2-acetylpyrrole was 0.91 ± 0.39 mg/kg and the OT in the source was 0.019 mg/kg (Lin et al., 2019); the ratio of concentration/OT value ≈ 48. In comparison, the average value of furaneol was 1.29 ± 0.73 mg/kg and the OT in sunflower oil was 0.025-0.05 mg/kg (van Gemert, 2011); the ratio of concentration/OT value ≈ 52. 2-Methoxy-4-vinylphenol (RI = 1787) is a naturally present phenolic compound which can be found in peanuts and clove (Jeong & Jeong, 2010) and commonly exist in peanut oil (Hu et al., 2014). The average value of 2-methoxy-4-vinylphenol was 2.03 ± 0.86 mg/kg and the OT in the oil was 0.05 mg/kg (van Gemert, 2011); the ratio of concentration/OT value ≈ 40. In addition, dibutyl phthalate (RI = 2290) existed in

Table 4.2. The univariate descriptive analysis of the quality traits of peanut butters (n = 40).

Traits	Average	Min	Max	SD	Q1	Q3
2,5-Dimethylpyrazine (mg/kg)	5.10	2.08	11.00	2.03	3.41	6.17
2,3,5-Trimethylpyrazine (mg/kg)	2.17	0.62	5.46	0.99	1.40	2.62
3-Ethyl-2,5-dimethylpyrazine (mg/kg)	1.41	0.57	3.66	0.66	0.91	1.79
2-Acetylpyrrole (mg/kg)	0.91	0.23	1.83	0.39	0.74	1.12
Furaneol (mg/kg)	1.29	0.15	2.89	0.73	0.59	1.78
2-Methoxy-4-vinylphenol (mg/kg)	2.03	0.39	4.77	0.86	1.45	2.39
Dibutyl phthalate (mg/kg)	0.78	0.08	3.93	0.87	0.27	0.95
L*	53.3	34.8	64.1	6.2	50.6	56.6
a*	9.6	3.5	13.5	2.7	8.8	11.8
b*	27.0	16.6	28.8	2.3	26.9	28.3
Firmness (g)	41.2	24.4	124.5	15.7	34.3	46.2
Consistency (g·s)	1058.0	646.7	3248.1	414.4	849.5	1203.7
Cohesiveness (g)	22.1	12.2	97.2	13.3	16.7	24.5
Index of viscosity (g·s)	389.5	63.1	2240.0	349.8	213.4	514.9
Yield stress (Pa)	9.13	2.79	26.28	4.14	6.21	11.20
Consistency coefficient (K, Pa·s^n)	5.24	2.19	22.74	3.57	3.28	5.57
Flow behaviour index (n)	0.82	0.65	0.88	0.05	0.80	0.86
G'- a (kPa Hz^{-b})	47.81	18.96	92.40	18.73	31.76	59.55
G'- b×100	18.35	12.85	28.89	3.45	15.83	20.23
G"- c (kPa Hz^{-d})	37.63	11.01	61.32	13.91	24.46	48.85
C"- d×100	34.10	21.45	46.08	4.53	31.37	36.60
D10 (μm)	8.14	5.30	10.83	1.32	7.23	8.68
D50 (μm)	48.70	28.06	70.26	12.91	39.58	61.15
D90 (μm)	151.95	89.25	192	21.60	141.75	163.40
Polydispersity	3.17	1.54	5.72	1.00	2.30	3.84

(D10) Average value of diameter below 10th percentile of particle size distributions; (D50) Average value of diameter below 50th percentile of particle size distributions; (D90) Average value of diameter below 90th percentile of particle size distributions; (*G*'- a) Power law coefficient of storage modulus; (*G*'- b) Power law exponent of storage modulus; (*G*"- c) Power law coefficient of loss modulus; (*G*"- d) Power law exponent of loss modulus; (Max) maximum; (Min) Minimum; (Q1) Quartile 1; (Q3) Quartile 3; (SD) Standard deviation.

roasting products (Akkad et al., 2019), where the average value was 0.78 ± 0.87 mg/kg and the OT in the air was 0.26 mg/L (van Gemert, 2011); the ratio of concentration/OT value ≈ 3.

The correlation analysis, as shown in **Figure 4.3b**, shows that there were significantly positive correlations ($P < 0.05$) between the seven volatile compounds except for dibutyl phthalate, which meant that peanut butter with aroma advantage had a higher content of the previously discussed aroma compounds. Among them, 2-acetylpyrrole was significantly positively correlated with pyrazines ($P < 0.05$). This was because 2-acetylpyrrole could be generated by the reaction of sucrose with other substances (Maarse, 2017). Furaneol and pyrazines are both aroma substances produced by peanuts during roasting (Liu et al., 2017). Therefore, there was a significantly positive correlation between them ($P < 0.05$).

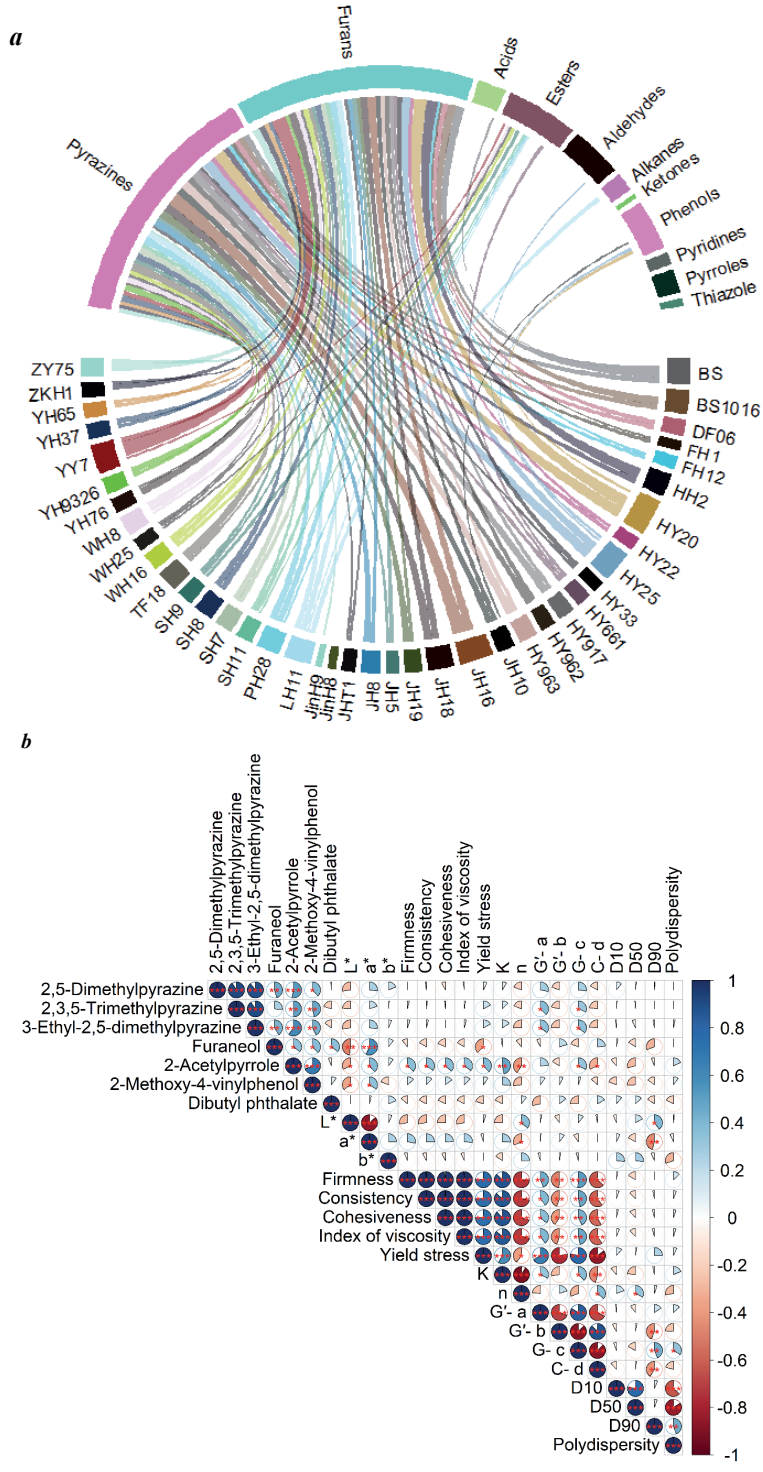
The colour is the most intuitive impression of peanut butter. The average colour values of the peanut butters were 53.3 ± 6.2 (L*), 9.6 ± 2.7 (a*), and 27.0 ± 2.3 (b*), respectively, which were also similar to the previous results (L*, 56; a*, 10; b*, 26) (Ha et al., 2013; Sanders et al., 2014). Among all peanut butters, JinH8 had a bright white colour with the highest L* (64.1 ± 0.1) and the lowest a* (3.5 ± 0.1) values. In comparison, ZY75 had a dark puce colour and the lowest L* (34.8 ± 0.1), the highest a* (12.8 ± 0.1), and the lowest b* (16.6 ± 0.1) values. The variance of colour between different varieties could be huge due to the qualities of raw materials under the same roasting conditions. More pictures are provided in **Figure S4.1**. There was a significantly negative relationship between L* and a* ($P < 0.05$) in **Figure 4.3b**. In other words, an increase in a* value increased the red value, resulting in a darker colour. In addition, L* had the significantly opposite trend with volatile compounds formed through reactions such as the Maillard reaction, especially 2-acetylpyrrole, furaneol, and 2-methoxy-4-vinylphenol ($P < 0.05$). Meanwhile, a higher roasting level produced a darker peanut butter (low L* value). Therefore, one can assume that peanut butters with lower L* values have higher concentrations of the above-mentioned volatiles.

The average texture values of the peanut butters were as follows: firmness, 41.2 ± 15.7 g; consistency, 1058.0 ± 414.4 g·s; cohesiveness, 22.1 ± 13.3 g; index of viscosity, 389.5 ± 349.8 g·s. The texture characteristics of different peanut butters were significantly varied (Appendix supplement can be found online at https://doi. org/10.1016/j.lwt.2021.112068, Appendix supplement 3). For example, JHT1 had the highest texture values (firmness, 124.5 ± 1.1 g; consistency, 3248.1 ± 34.6 g·s; cohesiveness, 97.2 ± 1.2 g; index of viscosity, 2240.0 ± 4.6 g·s), while YH76 had the lowest texture values (firmness, 24.4 ± 0.2 g; consistency, 646.7 ± 4.7 g·s; cohesiveness, 12.2 ± 0.1 g; index of viscosity, 63.1 ± 2.8 g·s). Firmness and consistency are reported to be related to the hardness of peanut butter, which are the sensory attributes evaluated at the initial and tongue-palate compression of oral processing (Koc et al., 2013; Nasaruddin et al., 2012). Cohesiveness and the index of viscosity are the tendency of peanut butter to cohere or stick together, while these properties also correlate with the slipperiness and adhesiveness of peanut butter (Koc et al., 2013). In the oral cavity, it is the speed at which peanut butter spreads on the tongue. The greater the cohesion, the slower the diffusion (Arrieta-Escobar et al., 2019). Therefore, the diversity levels of texture qualities were from YH76 to JHT1, which provides a basis reference for enterprises to select varieties. The correlation analysis (**Figure 4.3b**) shows that firmness and cohesiveness were significantly correlated ($P < 0.05$) with consistency and the index of viscosity. Hence, firmness and cohesiveness were chosen as the main texture characteristics for further analysis.

According to the results of the shear rate sweep, the viscosity of peanut butters decreased as the shear rate increased, reflecting that peanut butter is a typical non-New-

tonian fluid with pseudoplastic characteristics. The relationship between shear stress and shear rate could be calculated by Herschel-Bulkley. The fitting coefficients of all peanut butters were above 0.99. The models have three important parameters: yield stress (σ_0, Pa), K (consistency coefficient, Pa·s^n), and n (flow behavior index). The yield stress is the minimum shear stress needed to break the peanut butter's internal structure in order to initiate peanut butter flow (Augusto et al., 2012). If stress value is under the yield stress, peanut butter performs like an elastic solid. In contrast, it performs like a viscous liquid (Augusto et al., 2012). The average value of yield stress for the peanut butters was 9.13 ± 4.09 Pa. The difference among different varieties were obvious. JHT1 had the largest yield stress value (26.28 ± 2.53 Pa), making it the most difficult peanut butter to flow. On the contrary, the yield stress of YH76 was only 2.79 ± 0.11 Pa, which made it the easiest to start flowing. K is related to the energy to maintain flow and is directly proportional to the viscosity of the sample (Suhui et al., 2019). The average value of K was 5.24 ± 3.52 Pa·s^n. Among all peanut butters, JHT1 had the highest K value (22.74 ± 1.52 Pa·s^n) and resultingly the highest viscosity. The n value reflects flow behavior of peanut butter. The range of n value was from 0.65 to 0.88 which was less than one in all formulations showing non-Newtonian shear thinning behavior, which was similar to the previous results of shear rate sweep. For dynamic characteristics, storage modulus (*G*') presents the power reserved during every cycle of dynamic oscillation reflecting the elastic characteristics of peanut butter, while loss modulus (*G*") exposits the power dissipation in relation to the viscous characteristics (Wang, He, et al., 2017). All peanut butters showed gel characteristics because of *G*'> *G*" at all frequencies (Liu et al., 2019). The *G*'-a and *G*"-c of Power law model ($R^2 > 0.99$) indicate the amount of *G′* and *G′′* at a given frequency (Resch et al., 2002). The consistently higher *G*'-a than *G*"-c represented the aforesaid greater *G′* than *G′′* (Norazatul Hanim et al., 2016). The *G*'-b and *G*"-d reflect the magnitude of dependence of G′ and G′′ on frequency (Norazatul Hanim et al., 2016). The *G*'-b of all samples ranged from 12.85 - 28.89, which signified that peanut butter was a weak physical gel product (Bayod et al., 2008; Tunick, 2011). It could be found that n, *G*'-b, and *G*"-d reflected the same properties for all peanut butters, while Yield stress, K, *G*'-a, and *G*"-c varied greatly between different peanut butters.

According to Eq. (4), the polydispersity reflects the particle size distributions, which means the higher the value, the wider the particle size distributions. The average polydispersity value of the peanut butters was 3.17 ± 0.99, which is similar to that of commercial peanut butter (3.36 ± 0.08) (Norazatul Hanim et al., 2016). Among all peanut butters, FH12 had the lowest polydispersity value (1.54), while JinH8 had the highest value (5.72). With the decrease in particle size, the particle-particle interaction strength also decreases affecting the rheological characteristics and the mouthfeel of peanut butter in oral processing (Norazatul Hanim et al., 2016; Stokes et al., 2013). According to the **Figure 4.3b**, the polydispersity had a significantly positive relationship with *G*"-c ($P < 0.05$). This means that the decrease in particle size led to a decrease in viscosity.


a
Furans
Acids
Esters
Aldehydes
Alkanes
Ketones
Phenols
Pyridines
Pyrroles
Thiazole
Pyrazines
ZY75
ZKH1
YH65
YH37
YY7
YH326
YH76
WH8
WH25
WH16
TF18
SH9
SH8
SH7
SH11
PH28
LH11
JinH9
JinH8
JHT1
JH8
JH5
JH19
JH18
JH16
JH10
HY963
HY962
HY917
HY661
HY33
HY25
HY22
HY20
HH2
FH12
FH1
DF06
BS1016
BS
b
2,5-Dimethylpyrazine
2,3,5-Trimethylpyrazine
3-Ethyl-2,5-dimethylpyrazine
Furaneol
2-Acetylpyrrole
2-Methoxy-4-vinylphenol
Dibutyl phthalate
L*
a*
b*
Firmness
Consistency
Cohesiveness
Index of viscosity
Yield stress
K
n
G'- a
G'- b
G- c
C- d
D10
D50
D90
Polydispersity
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
-1

4.3.4. The traits of the peanut butters: multivariate comparisons

K-means clustering combined with PCA was applied to elaborate the relationships between the different peanut butters. To balance the weights of different variables, raw data was auto-scaled before PCA. JHT1 with extreme values of textural and rheological properties was removed because it was located outside the ellipse of the Hotelling's T^2 (0.5%) (Yu, Liu, Wang, et al., 2020). An overview of the differences between the peanut butters relating to all tested parameters is shown in **Figure 4.4a**. The first two PCs of the model explained 52% of the total variance with a contribution of 32% from PC1. In addition to dibutyl phthalate, b*, and polydispersity, other variables mainly contributed to PC1 and PC2. Among them, pyrazines, 2-acetylpyrrole, firmness, and K had the greatest contributions to PC1, while furaneol, yield stress, a*, L* and *G*"-c had the greatest contributions to PC2. According to the K-means results, three clusters of peanut butters could be identified on the PCA biplot. Cluster 1 (HY25, JH16, HY963, etc.) grouped on the right of the PCA biplot (the opposite direction of PC1). Most parameters had the same direction with this group, such as volatile compounds, texture, and rheological qualities. Cluster 2 (SH9, YH65, JinH8, etc.) located on the top left of the PCA biplot (the negative direction of PC1 and the positive direction of PC2). This group presented a positive relationship with L* and negative relationship with volatile compounds. Cluster 3 (BS1016, HY33, YY7, etc.), situated on the bottom left of the PCA biplot. This cluster associated positively with dibutyl phthalate and negatively with texture and rheological qualities.

The results of the different parameters in different peanut butter groups are shown in **Figure 4.4b**. Most parameters had significant differences between different groups except for 2-methoxy-4-vinylphenol, dibutyl phthalate, b*, and polydispersity. Each group had inherent characteristics. Specifically, group 1 had significantly higher contents of most volatile compounds than the other groups, especially pyrazines ($P < 0.05$). Group 2 was significantly different from the other two groups for colour (L* and a* values) ($P < 0.05$). The values for firmness and cohesiveness of group 1 and group 2 were significantly higher than that of group 3. Similar results could also be found for rheology

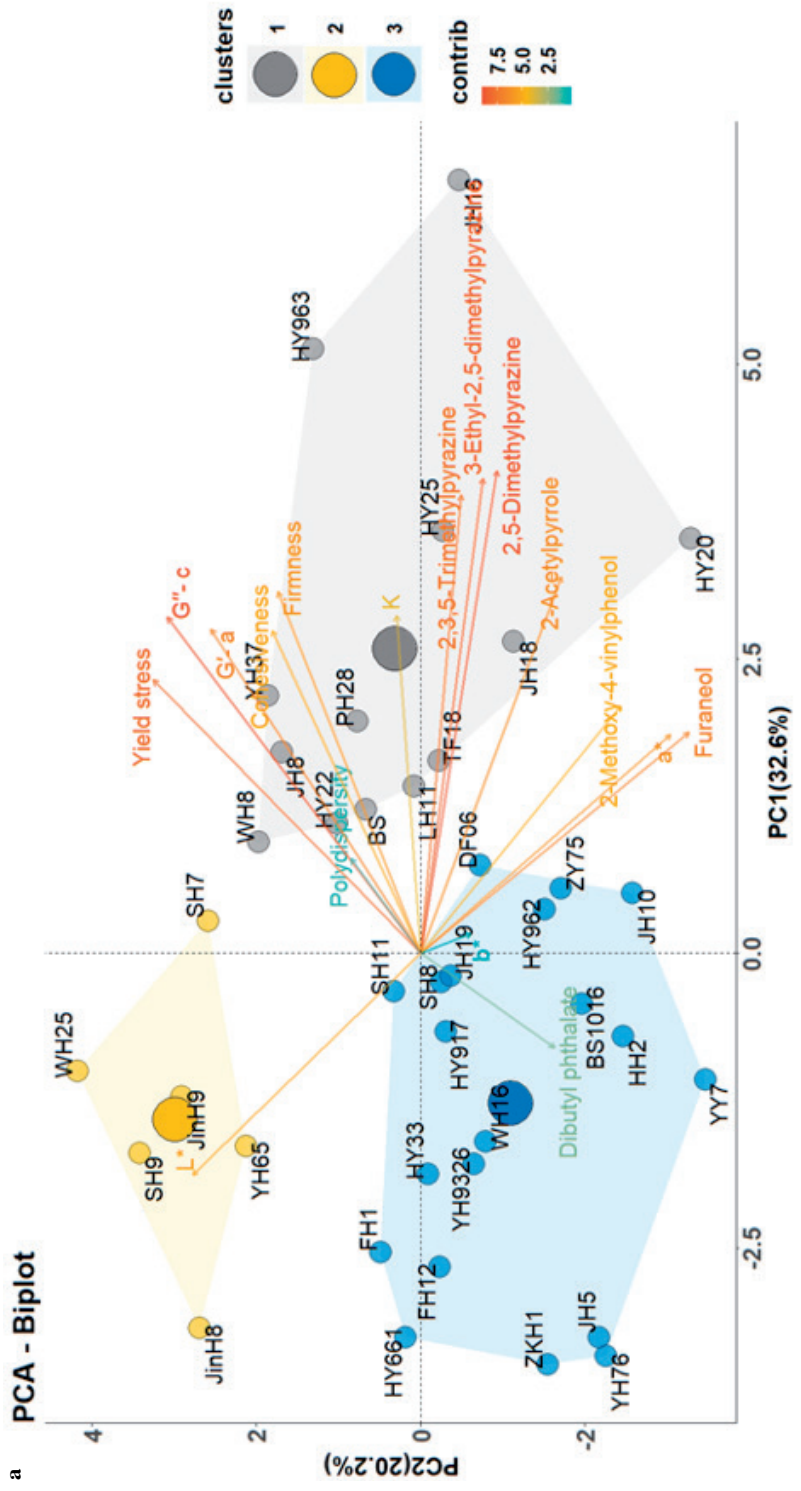
◀ **Figure 4.3.** The (a) volatile compounds of the different peanut butters (the link lines indicate that the contents of volatile compounds that exceeded 5 mg/kg), and the (b) correlation analysis of the qualities of the peanut butters. The positive and negative correlation coefficients (R) are coloured blue and red, respectively. The fraction of the circles stands for the correlation coefficient and full fraction means R = 1. (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (D10) Average value of diameter below 10th percentile of particle size distributions; (D50) Average value of diameter below 50th percentile of particle size distributions; (D90) Average value of diameter below 90th percentile of particle size distributions; (*G*'- a) Power law coefficient of storage modulus; (*G*'- b) Power law exponent of storage modulus; (*G*"- c) Power law coefficient of loss modulus; (*G*"- d) Power law exponent of loss modulus; (K) Consistency coefficient; (n) Flow behaviour index.

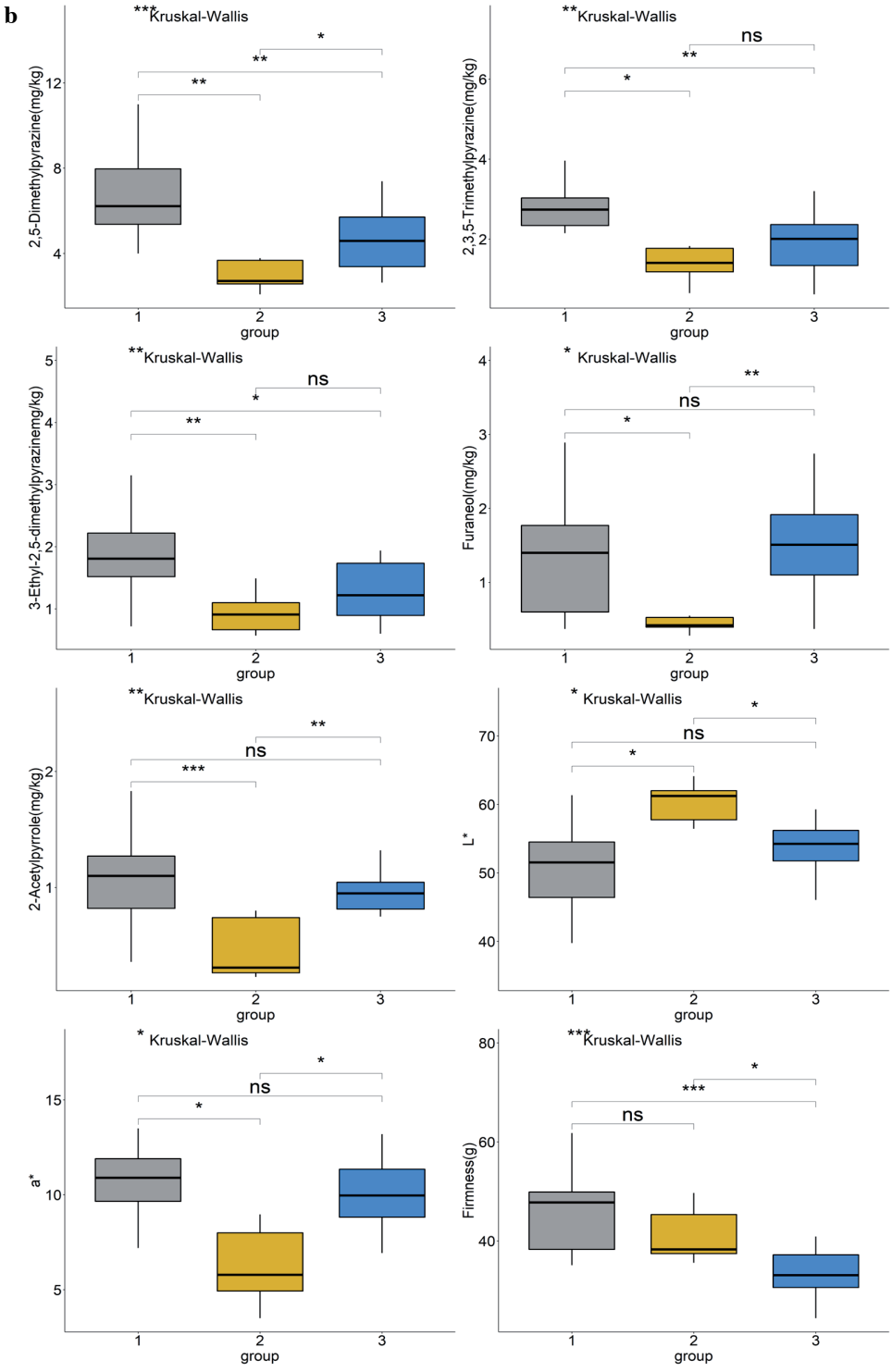
($P < 0.05$). In summary, group 1 had the highest pyrazine contents, the highest values of texture and rheology qualities, and a moderate level of colour.

4.3.5. Relationships between peanut varieties and peanut butters

The correlation analysis results of the quality traits of the peanut varieties versus all parameters of the peanut butters are shown in **Figure 4.5a**. In terms of the effects of the raw materials, the texture, rheology, and polydispersity were generally bracketed, mainly reflecting the internal structure of peanut butter, while the colour and aroma compounds were correspondingly clustered together. Specifically, sucrose had a significantly positive correlation with the texture (firmness and cohesiveness) and the rheology (yield stress, K, *G'*-a, and *G''*-c) of the peanut butters. The Pearson correlation coefficient (R) index between sucrose and the above-mentioned parameters were: $R_{firmness}$ (0.60, $P < 0.05$), $R_{cohensiveness}$ (0.58, $P < 0.05$), $R_{yield\ stress}$ (0.49, $P < 0.05$), R_K (0.56, $P < 0.05$), $R_{G'\text{-}a}$ (0.36, $P < 0.05$), and $R_{G''\text{-}c}$ (0.47, $P < 0.05$) respectively. This could be since sucrose is known to lead to an increase in the macromolecular entanglement (Mathlouthi & Reiser, 2012; Wang et al., 2009). As a result, the corresponding R indexes of texture and rheology were increased. In addition, the peanut varieties with higher sucrose contents always had a lower fat content and a higher protein content (Bishi et al., 2013). Importantly, the fat content had a significantly negative relationship with the texture and rheological characteristics such as $R_{firmness}$ (-0.39, $P < 0.05$), while the protein content had a significantly positive relationship with these parameters such as $R_{firmness}$ (0.32, $P < 0.05$). This is because the continuous fat phase acts as a "lubricant" in the suspension system of peanut butter, while the macromolecules like protein are distributed among it (Citerne et al., 2001; Mohd Rozalli et al., 2016). Once the fat phase is reduced or the macromolecular substance is increased, the viscosity will increase. Meanwhile, protein as the controlling trait for a stronger gel network makes the samples firmer (Nguyen et al., 2017). Among the measured amino acids, proline (Pro) and Cys were significantly positively correlated with the texture and rheological characteristics. The R indexs between Pro and the following parameters were: $R_{firmness}$ (0.55, $P < 0.05$), $R_{cohensiveness}$ (0.54, $P < 0.05$), and $R_{yield\ stress}$ (0.44, $P < 0.05$). Similarly, Cys was postively correlated with the following parameters: $R_{firmness}$ (0.34, $P < 0.05$) and $R_{cohensiveness}$ (0.37, $P < 0.05$). The reason is that Pro can enhance the hydrophobicity of the internal structure of protein to improve the stability (Levitt, 1978), while Cys can form disulfide bonds (Gudmundsson, 2002). All of these traits could increase the stability of the protein space structure, resulting in the strength of the system.

Amino acids (nitrogen supply) and saccharides (carbon sources) are two important precursor substances known to contribute to the formation of pyrazines (Hui et al., 2010). For instanse, Ser decomposition products could react with sucrose to form pyrazines (Hui et al., 2010). Ser had a significantly postive relationship with 2,5-dimethylpyrazine and


a
PCA - Biplot
clusters
1
2
3
contrib
7.5
5.0
2.5
PC1(32.6%)
PC2(20.2%)
5.0
2.5
0.0
-2.5
4
2
0
-2
Yield stress
G"- c
G'- a
Consistency
Firmness
K
Polydispersity
2,3,5-Trimethylpyrazine
3-Ethyl-2,5-dimethylpyrazine
2,5-Dimethylpyrazine
2-Acetylpyrrole
2-Methoxy-4-vinylphenol
Furaneol
L*
a*
b*
Dibutyl phthalate
HY963
HY25
HY20
JH16
JH18
PH28
YH37
JH8
HY22
WH8
BS
LH11
TF18
DF06
ZY75
HY962
JH10
SH7
SH11
SH8
JH19
WH25
JinH9
SH9
YH65
JinH8
HY917
BS1016
HH2
YY7
WH16
HY33
YH9326
FH1
FH12
HY661
ZKH1
JH5
YH76


b
***Kruskal-Wallis
2,5-Dimethylpyrazine(mg/kg)
**Kruskal-Wallis
2,3,5-Trimethylpyrazine(mg/kg)
**Kruskal-Wallis
3-Ethyl-2,5-dimethylpyrazinemg/kg)
*Kruskal-Wallis
Furaneol(mg/kg)
**Kruskal-Wallis
2-Acetylpyrrole(mg/kg)
*Kruskal-Wallis
L*
*Kruskal-Wallis
a*
***Kruskal-Wallis
Firmness(g)
group
ns

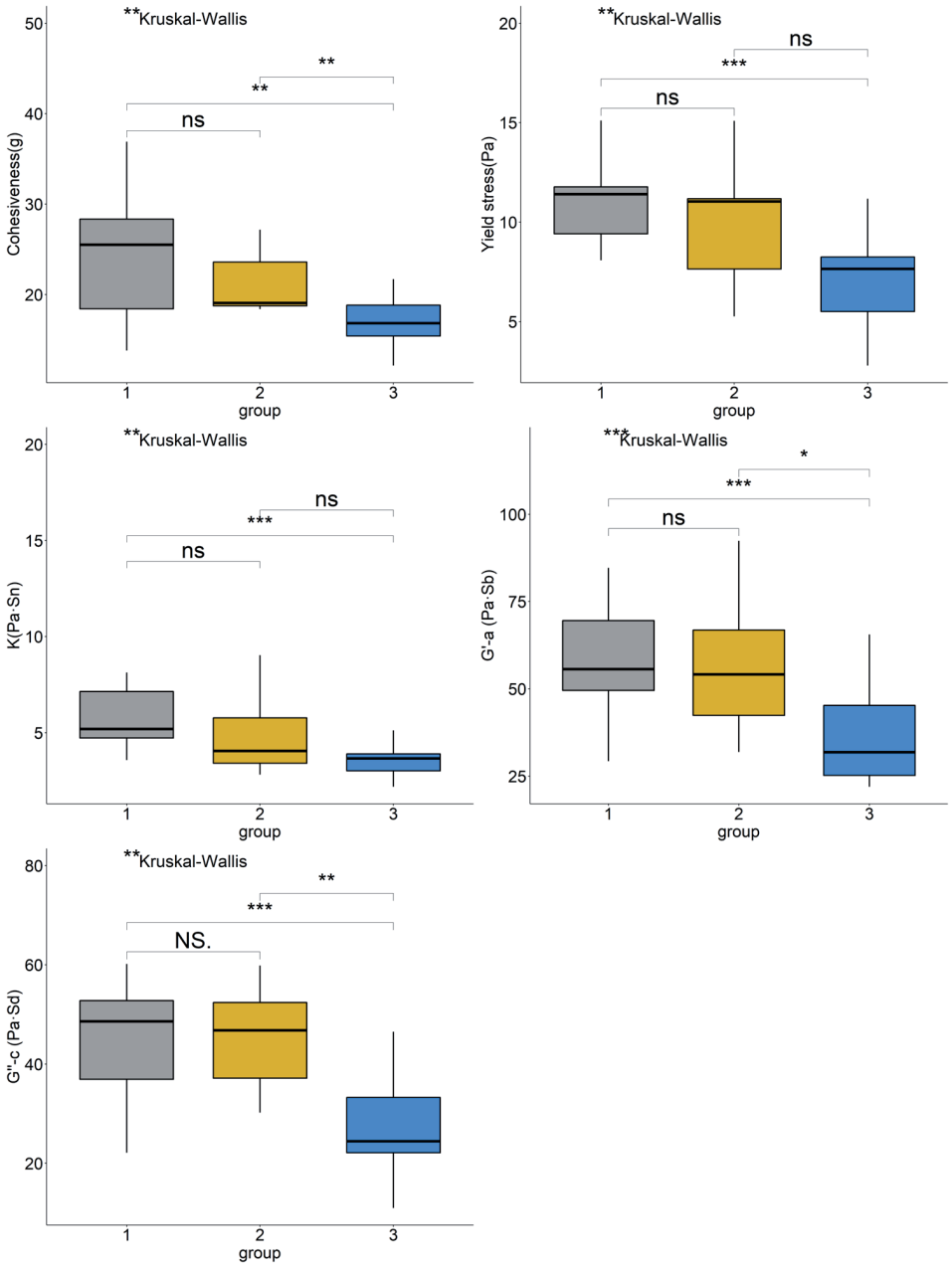


Figure 4.4. The (a) principal component analysis (PCA) biplot of the different peanut butters (n = 39). The different colored polygons indicate the different sample clusters and arrows indicate the contributions of parameters), and the (b) significantly different qualities of peanut butters in different groups (NS. $P = 1$, ns $1 > P >= 0.05$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$). (contrib) Contribution; (*G*'- a) Power law coefficient of storage modulus; (*G*"- c) Power law coefficient of loss modulus; (K) Consistency coefficient.

3-ethyl-2,5-dimethylpyrazine ($P < 0.05$) (**Figure 4.5a**). Meanwhile, Ala, histidine (His), and Leu can react with monosaccharides to produce pyrazines as well (Hui et al., 2010). Therefore, these amino acids likely had a positive contribution to the formation of pyrazines in the peanuts. As a nitrogen-containing heterocyclic compounds, a higher concentration of 2-acetylpyrrole have been observed with the increase in sucrose content (Saputro et al., 2018). Correspondingly, the relationship between 2-acetylpyrrole and sucrose was significantly positive (0.36, $P < 0.05$) in this study. On the contrary, fat had a significantly negatively influence on 2-acetylpyrrole formation (-0.38, $P < 0.05$) as it could have reacted with lipid degradation products (Zhao, Wang, et al., 2019). It has been found that dibutyl phthalate as fat-soluble substance can increase if the fat content increases (Cao, 2010). This could explain its significantly positive relationship with fat (0.32, $P < 0.05$) as shown in **Figure 4.5a**. Glycine (Gly) and monosaccharides could react with 2-methoxy-4-vinylphenol to produce an increase in furfural (Jiang et al., 2009), resulting in a significant negative relationship between Gly and 2-methoxy-4-vinylphenol (-0.37, $P < 0.05$). The colours are affected by the peanut size as the surface area of large peanut kernels is smaller than those of small kernels at the same weight during roasting. The weight is calculated based on 100 kernels per variety and the higher the weight, the smaller the surface area. As a result, large peanut kernels were roasted to a lower extent, leading to higher L* values for the peanut butters at the same roasting conditions. Therefore, L* had a significant positive relationship with weight (0.32, $P < 0.05$) as shown in **Figure 4.5a**.

Figure 4.5b shows the correlation between the quality traits of the peanut varieties and the score matrixes from the PCA of the peanut butters in section 4.3.4. The first five components were regarded as the effective components because their eigenvalues were higher than one (Wang et al., 2020). PC1 had a significant negative relationship with fat (-0.40, $P < 0.05$) and a significant positive relationship with sucrose (0.32, $P < 0.05$). This shows that peanut butters processed from the peanut varieties that contained higher sucrose and lower fat contents was situated on the right side of **Figure 4.4a**, corresponding to the positive direction of PC1 which is associated with the contributions from texture, rheology, and pyrazines. The positive direction of PC2 stem from texture, rheology, and L*, while the negative direction of PC2 is attributed to the most volatile compounds (**Figure 4.4a**). PC2 had significant positive relationships with Tyr (0.40, $P < 0.05$), Leu (0.38, $P < 0.05$), and Thr (0.32, $P < 0.05$). In other words, peanut butters processed from the peanut varieties with higher contents of the above-mentioned amino acids normally located on the top side of **Figure 4.4a** (the positive direction of PC2).

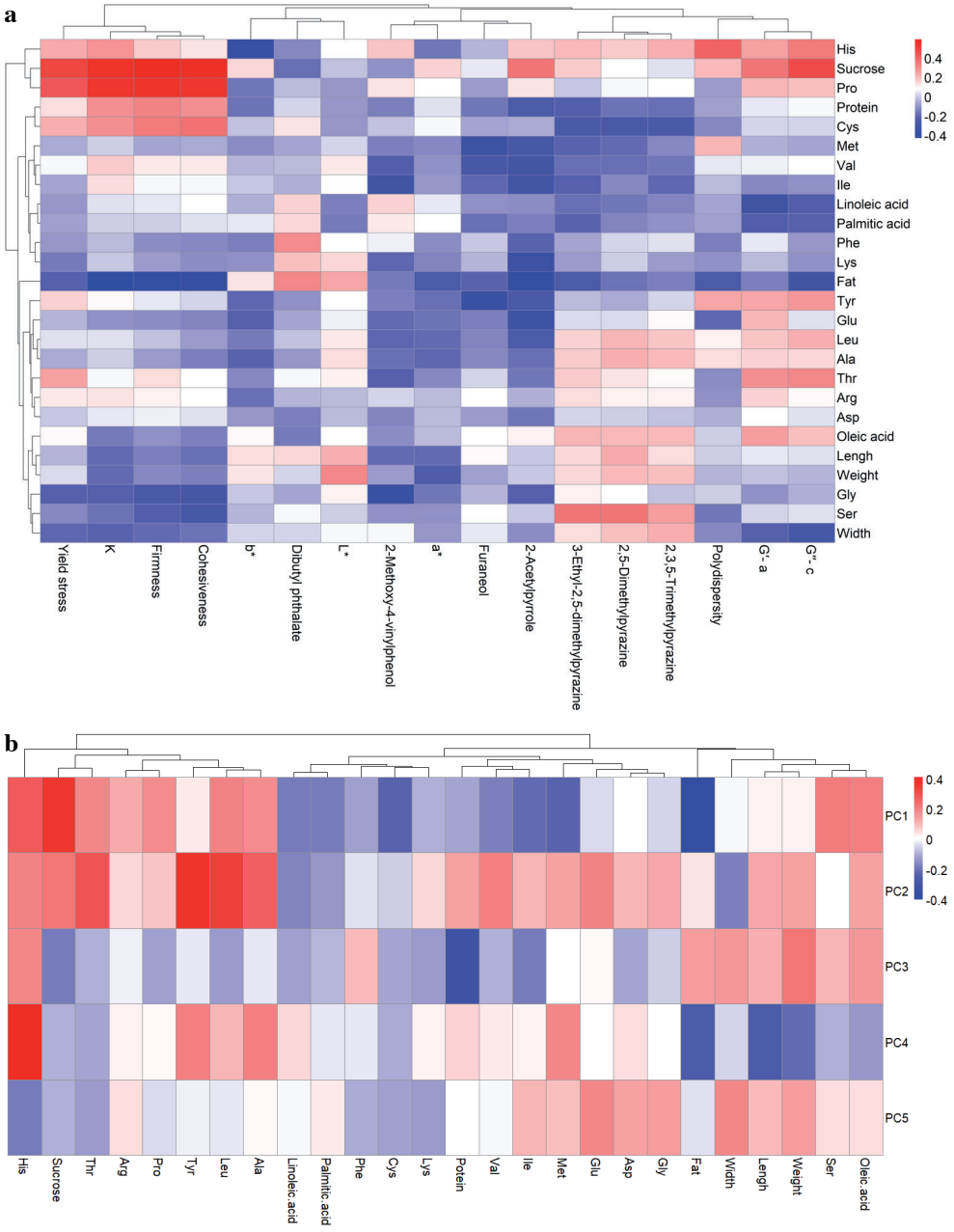


Figure 4.5. The (a) correlation analysis between quality traits of the peanut varieties (vertical axis) and qualities of the peanut butter (horizontal axis), and the (b) correlation analysis between the first five principal components of the peanut butters (vertical axis) and quality traits of the peanut varieties (horizontal axis). The positive and negative coefficients are coloured red and blue, respectively. Coefficients < -0.32 and > + 0.32 indicate significant correlations ($P < 0.05$). (Ala) Alanine; (Arg) Arginine; (Asp) Aspartic acid; (Cys) Cysteine; (His) Histidine; (Ile) Isoleucine; (Glu) Glutamic acid; (Gly) Glycine; (Leu) Leucine; (Lys) Lysine; (Met) Methionine; (Phe) Phenylalanine; (Pro) Proline; (Ser) Serine; (Thr) Threonine; (Tyr) Tyrosine; (Val) Valine.

4.4. CONCLUSIONS AND IMPLICATIONS

A wide diversity in quality traits could be found among the forty peanut varieties. Thereby many heterogeneities in the properties of the peanut butters were discovered based on their volatile compounds, colour, texture, rheology, and particle size distributions. PCA with K-means analysis showed different regions with homogeneous characteristics of the peanut butters. Among them, group 1 (HY 25, JH 18, YH37, etc.) had the highest pyrazines contents and the highest values for texture and rheological properties. In general, it was proposed that the peanut varieties with lower fat, higher protein and sucrose contents could be used to manufacture peanut butters with higher values of the textural and rheological qualities studied; whereas, the peanut varieties with higher sucrose and Ser contents could be used to manufacture peanut butters with higher concentrations of pyrazines and 2-acetylpyrrole, while the heavier peanut varieties with higher Tyr and Thr contents could be used to manufacture peanut butters with higher values of L* and rheological qualities. Instead of focusing on one or two properties from a small data set, this study considered the most important properties of peanut butters that are needed for selecting peanut varieties. This approach will better assist processing enterprises to select the most suitable peanut varieties to produce peanut butters for different commercial needs.

REFERENCES

Akkad, R., Kharraz, E., Han, J., House, J. D., & Curtis, J. M. (2019). Characterisation of the volatile flavour compounds in low and high tannin faba beans (Vicia faba var. minor) grown in Alberta, Canada. *Food Research International, 120*, 285-294. https://doi.org/10.1016/j.foodres.2019.02.044.

Arrieta-Escobar, J. A., Bernardo, F. P., Orjuela, A., Camargo, M., & Morel, L. (2019). Incorporation of heuristic knowledge in the optimal design of formulated products: Application to a cosmetic emulsion. *Computers & Chemical Engineering, 122*, 265-274. https://doi.org/10.1016/j.compchemeng.2018.08.032.

Arsa, S., & Theerakulkait, C. (2018). Preparation, aroma characteristics and volatile compounds of flavorings from enzymatic hydrolyzed rice bran protein concentrate. *Journal of the Science of Food and Agriculture, 98*(12), 4479-4487. https://doi.org/10.1002/jsfa.8972.

Augusto, P. E. D., Cristianini, M., & Ibarz, A. (2012). Effect of temperature on dynamic and steady-state shear rheological properties of siriguela (Spondias purpurea L.) pulp. *Journal of Food Engineering, 108*(2), 283-289. https://doi.org/10.1016/j.jfoodeng.2011.08.015.

Baker, G., Cornell, J., Gorbet, D., O'keefe, S., Sims, C., & Talcott, S. (2003). Determination of pyrazine and flavor variations in peanut genotypes during roasting. *Journal of Food Science, 68*(1), 394-400. https://doi.org/10.1111/j.1365-2621.2003.tb14171.x.

Bayod, E., Willers, E. P., & Tornberg, E. (2008). Rheological and structural characterization of tomato paste and its influence on the quality of ketchup. *Lwt-Food Science and Technology, 41*(7), 1289-1300. https://doi.org/10.1016/j.lwt.2007.08.011.

Bishi, S. K., Kumar, L., Dagla, M. C., Mahatma, M. K., Rathnakumar, A. L., Lalwani, H. B., & Misra, J. B. (2013). Characterization of Spanish peanut germplasm (Arachis hypogaea L.) for sugar profiling and oil quality. *Industrial Crops and Products, 51*, 46-50. https://doi.org/10.1016/j.indcrop.2013.08.050.

Cao, X. L. (2010). Phthalate esters in foods: sources, occurrence, and analytical methods. *Comprehensive reviews in food science and food safety, 9*(1), 21-43. https://doi.org/10.1111/j.1541-4337.2009.00093.x.

Chen, B., Li, Q., Hu, H., Meng, S., Shah, F., Wang, Q., & Liu, H. (2020). An optimized industry processing technology of peanut tofu and the novel prediction model for suitable peanut varieties. *Journal of Integrative Agriculture, 19*(9), 2340-2351. https://doi.org/10.1016/s2095-3119(20)63249-x.

Citerne, C. P., Pierna, J. C., & Moan, M. (2001). Rheological properties of peanut butter. *Rheologica Acta, 40*, 86-96. https://doi.org/10.1007/s003970000120.

Dhamsaniya, N. K., Patel, N. C., & Dabhi, M. N. (2012). Selection of groundnut variety for making a good quality peanut butter. *Journal of Food Science and Technology, 49*(1), 115-118. https://doi.org/10.1007/s13197-011-0361-9.

GB5009.5. (2016). *National Food Safety Standard - Determination of Protein in Foods*. Beijing: Standard Press of China.

GB5009.6. (2016). *National Food Safety Standard - Determination of Fat in Foods*. Beijing: Standard Press of China.

GB5009.8. (2016). *National food safety standard - Determination of fructose, glucose, sucrose, maltose and lactose in foods* Beijing: Standard Press of China.

GB5009.124. (2016). *National Food Safety Standard - Determination of Amino Acid in Foods*. Beijing: Standard Press of China.

GB5009.168. (2016). National food safety standard- Determination of Fatty Acids in Foods *Beijing: Standard Press of China*.

Gong, A., Shi, A., Liu, H., Yu, H., Liu, L., Lin, W., & Wang, Q. (2018). Relationship of chemical properties of different peanut varieties to peanut butter storage stability. *Journal of Integrative Agriculture, 17*(5), 1003-1010. https://doi.org/10.1016/s2095-3119(18)61919-7.

Gudmundsson, M. (2002). Rheological properties of fish gelatins. *Journal of Food Science, 67*(6), 2172-2176. https://doi.org/10.1111/j.1365-2621.2002.tb09522.x.

Ha, J. W., Kim, S. Y., Ryu, S. R., & Kang, D. H. (2013). Inactivation of Salmonella enterica serovar Typhimurium and Escherichia coli O157:H7 in peanut butter cracker sandwiches by radio-frequency heating. *Food Microbiol, 34*(1), 145-150. https://doi.org/10.1016/j.fm.2012.11.018.

Hashemian, M., Murphy, G., Etemadi, A., Dawsey, S. M., Liao, L. M., & Abnet, C. C. (2017). Nut and peanut butter consumption and the risk of esophageal and gastric cancer subtypes. *The American Journal of Clinical Nutrition, 106*(3), 858-864. https://doi.org/10.3945/ajcn.117.159467.

Hu, W., Zhang, L., Li, P., Wang, X., Zhang, Q., Xu, B., Sun, X., Ma, F., & Ding, X. (2014). Characterization of volatile components in four vegetable oils by headspace two-dimensional comprehensive chromatography time-of-flight mass spectrometry. *Talanta, 129*, 629-635. https://doi.org/10.1016/j.talanta.2014.06.010.

Hui, Y. H., Chen, F., Nollet, L. M., Guiné, R. P., Martín-Belloso, O., Mínguez-Mosquera, M. I., Paliyath, G., Pessoa, F. L., Le Quéré, J.-L., & Sidhu, J. S. (2010). *Handbook of fruit and vegetable flavors*. Hoboken: John Wiley and Sons.

Jeong, J. B., & Jeong, H. J. (2010). 2-Methoxy-4-vinylphenol can induce cell cycle arrest by blocking the hyper-phosphorylation of retinoblastoma protein in benzo[a]pyrene-treated NIH3T3 cells. *Biochemical and Biophysical Research Communications, 400*(4), 752-757. https://doi.org/10.1016/j.bbrc.2010.08.142.

Jiang, D., Chiaro, C., Maddali, P., Prabhu, K. S., & Peterson, D. G. (2009). Identification of Hydroxycinnamic Acid–Maillard Reaction Products in Low-Moisture Baking Model Systems. *Journal of Agricultural and Food Chemistry, 57*(21), 9932-9943. https://doi.org/10.1021/jf900932h.

Jiang, H., Duan, N., & Ren, X. (2006). *Specification of peanut germplasm resources*. Beijing: China Argricultrue Press.

Klevorn, C. M., Dean, L. L., & Johanningsmeier, S. D. (2019). Metabolite Profiles of Raw Peanut Seeds Reveal Differences between Market-Types. *Journal of Food Science, 84*(3), 397-405. https://doi.org/10.1111/1750-3841.14450.

Koc, H., Vinyard, C. J., Essick, G. K., & Foegeding, E. A. (2013). Food oral processing: conversion of food structure to textural perception. *Annual Review of Food Science and Technology, 4*, 237-266. https://doi.org/10.1146/annurev-food-030212-182637.

Levitt, M. (1978). Conformational preferences of amino acids in globular proteins. *Biochemistry, 17*(20), 4277-4285. https://doi.org/10.1021/bi00613a026.

Li, C., & Hou, L. (2018). Review on Volatile Flavor Components of Roasted Oilseeds and Their Products. *Grain & Oil Science and Technology, 1*(4), 151-156. https://doi.org/10.3724/sp.J.1447.Gost.2018.18052.

Lin, H., Liu, Y., He, Q., Liu, P., Che, Z., Wang, X., & Huang, J. (2019). Characterization of odor components of Pixian Douban (broad bean paste) by aroma extract dilute analysis and odor activity values. *International Journal of Food Properties, 22*(1), 1223-1234. https://doi.org/10.1080/10942912.2019.1636816.

Liu, X., Jin, Q., Liu, Y., Huang, J., Wang, X., Mao, W., & Wang, S. (2011). Changes in volatile compounds of peanut oil during the roasting process for production of aromatic roasted peanut oil. *Journal of Food Science 76*(3), 404-412. https://doi.org/10.1111/j.1750-3841.2011.02073.x.

Liu, Y., Hu, H., Liu, H., Li, J., Yang, Y., Shi, A., Liu, L., & Wang, Q. (2017). Analysis of Volatile Compounds in Roasted Peanut Shell and Peanut Cakes by HS-SPME-GC-MS. *Food Science 2*, 139-145.

Liu, Y., Yu, Y., Liu, C., Regenstein, J. M., Liu, X., & Zhou, P. (2019). Rheological and mechanical behavior of milk protein composite gel for extrusion-based 3D food printing. *LWT - Food Science and Technology, 102*, 338-346. https://doi.org/10.1016/j.lwt.2018.12.053.

Lou, F., Liu, Y., Sun, X., Pan, Y., Zhao, J., & Zhao, Y. (2009). Identification of Volatile Flavor Components in Peanut Butter *Food Science, 24*, 393-396.

Maarse, H. (2017). *Volatile Compounds in Foods and Beverages*. Boca Raton: CRC Press.

Mathlouthi, M., & Reiser, P. (2012). *Sucrose: Properties and Applications*. New York: Springer

Mohd Rozalli, N. H., Chin, N. L., & Yusof, Y. A. (2015). Particle Size Distribution of Natural Peanut Butter and Its Dynamic Rheological Properties. *International Journal of Food Properties, 18*(9), 1888-1894. https://doi.org/10.1080/10942912.2014.971184.

Mohd Rozalli, N. H., Chin, N. L., Yusof, Y. A., & Mahyudin, N. (2016). Quality changes of stabilizer-free natural peanut butter during storage. *Journal of Food Science and Technology, 53*(1), 694-702. https://doi.org/10.1007/s13197-015-2006-x.

Nasaruddin, F., Chin, N. L., & Yusof, Y. A. (2012). Effect of Processing on Instrumental Textural Properties of TraditionalDodolUsing Back Extrusion. *International Journal of Food Properties, 15*(3), 495-506. https://doi.org/10.1080/10942912.2010.491932.

Nayak, S. N., Hebbal, V., Bharati, P., Nadaf, H. L., Naidu, G. K., & Bhat, R. S. (2020). Profiling of Nutraceuticals and Proximates in Peanut Genotypes Differing for Seed Coat Color and Seed Size. *Frontiers in Nutrition, 7*(45). https://doi.org/10.3389/fnut.2020.00045.

Nguyen, P. T. M., Kravchuk, O., Bhandari, B., & Prakash, S. (2017). Effect of different hydrocolloids on texture, rheology, tribology and sensory perception of texture and mouthfeel of low-fat pot-set yoghurt. *Food Hydrocolloids, 72*, 90-104. https://doi.org/10.1016/j.foodhyd.2017.05.035.

Norazatul Hanim, M. R., Chin, N. L., & Yusof, Y. A. (2016). Effects of Grinding Time on Rheological, Textural and Physical Properties of Natural Peanut Butter Stored at Different Temperatures. *Journal of Texture Studies, 47*(2), 131-141. https://doi.org/10.1111/jtxs.12167.

Resch, J. J., & Daubert, C. R. (2002). Rheological and Physicochemical Properties of Derivatized Whey Protein Concentrate Powders. *International Journal of Food Properties, 5*(2), 419-434. https://doi.org/10.1081/jfp-120005795.

Sanders, C. T., DeMasie, C. L., Kerr, W. L., Hargrove, J. L., Pegg, R. B., & Swanson, R. B. (2014). Peanut skins-fortified peanut butters: Effects on consumer acceptability and quality characteristics. *LWT - Food Science and Technology, 59*(1), 222-228. https://doi.org/10.1016/j.lwt.2014.04.001.

Saputro, A. D., Van de Walle, D., Hinneh, M., Van Durme, J., & Dewettinck, K. (2018). Aroma profile and appearance of dark chocolate formulated with palm sugar–sucrose blends. *European Food Research and Technology, 244*(7), 1281-1292. https://doi.org/10.1007/s00217-018-3043-2.

Shakerardekani, A., Karim, R., Ghazali, H. M., & Chin, N. L. (2013). Textural, rheological and sensory properties and oxidative stability of nut spreads–a review. *International Journal of Molecular Sciences, 14*(2), 4223-4241. https://doi.org/10.3390/ijms14024223.

Starowicz, M., & Zieliński, H. (2019). How Maillard Reaction Influences Sensorial Properties (Color, Flavor and Texture) of Food Products? *Food Reviews International, 35*(8), 707-725. https://doi.org/10.1080/87559129.2019.1600538.

Stokes, J. R., Boehm, M. W., & Baier, S. K. (2013). Oral processing, texture and mouthfeel: From rheology to tribology and beyond. *Current Opinion in Colloid & Interface Science, 18*(4), 349-359. https://doi.org/10.1016/j.cocis.2013.04.010.

Suhui, L., Xuede, W., Yuxiang, M., Jingyuan, L., & Guozhan, Z. (2019). Preparation of sesame peanut paste and its rheological properties. *China Oils and Fats, 02*, 147-152.

Taghizadeh, M., & Razavi, S. M. A. (2009). Modeling Time-Independent Rheological Behavior of Pistachio Butter. *International Journal of Food Properties, 12*(2), 331-340. https://doi.org/10.1080/10942910701772048.

Tamura, H., Fujita, A., Steinhaus, M., Takahisa, E., Watanabe, H., & Schieberle, P. (2010). Identification of Novel Aroma-Active Thiols in Pan-Roasted White Sesame Seeds. *Journal of Agricultural and Food Chemistry, 58*(12), 7368-7375. https://doi.org/10.1021/jf100623a.

Tunick, M. H. (2011). Small-strain dynamic rheology of food protein networks. *Journal of Agricultural and Food Chemistry, 59*(5), 1481-1486. https://pubs.acs.org/doi/pdf/10.1021/jf1016237.

van Gemert, L. J. (2011). *Odour thresholds: compilations of odour threshold values in air, water and other media*. Utrecht: Oliemans Punter & Partners.

Wang, B., Wang, L.-J., Li, D., Özkan, N., Li, S.-J., & Mao, Z.-H. (2009). Rheological properties of waxy maize starch and xanthan gum mixtures in the presence of sucrose. *Carbohydrate Polymers, 77*(3), 472-481. https://doi.org/10.1016/j.carbpol.2009.01.017.

Wang, F., Meng, J., Sun, L., Weng, Z., Fang, Y., Tang, X., Zhao, T., & Shen, X. (2020). Study on the tofu quality evaluation method and the establishment of a model for suitable soybean varieties for Chinese traditional tofu processing. *Lwt-Food Science and Technology, 117*. https://doi.org/10.1016/j.lwt.2019.108441.

Wang, L., Liu, H., Liu, L., Wang, Q., Li, S., & Li, Q. (2017). Prediction of peanut protein solubility based on the evaluation model established by supervised principal component regression. *Food Chemistry, 218*, 553-560. https://doi.org/10.1016/j.foodchem.2016.09.091.

Wang, L., Wang, Q., Liu, H., Liu, L., & Du, Y. (2013). Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy. *Journal of Food Science and Technology, 93*(1), 118-124. https://doi.org/10.1002/jsfa.5738.

Wang, Q. (2016). *Peanuts: Processing technology and product development*. Cambridge: Academic Press.

Wang, Q. (2018). *Peanut processing characteristics and quality evaluation*. New York: Springer.

Wang, Q., Liu, H., Shi, A., Hu, H., Liu, L., Wang, L., & Yu, H. (2017). Review on the processing characteristics of cereals and oilseeds and their processing suitability evaluation technology. *Journal of Integrative Agriculture, 16*(12), 2886-2897. https://doi.org/10.1016/s2095-3119(17)61799-4.

Wang, S., Adhikari, K., & Hung, Y. C. (2017). Acceptability and Preference Drivers of Freshly Roasted Peanuts. *Journal of Food Science 82*(1), 174-184. https://doi.org/10.1111/1750-3841.13561.

Wang, X., He, Z., Zeng, M., Qin, F., Adhikari, B., & Chen, J. (2017). Effects of the size and content of protein aggregates on the rheological and structural properties of soy protein isolate emulsion gels induced by CaSO4. *Food Chemistry, 221*, 130-138. https://doi.org/10.1016/j.foodchem.2016.10.019.

Yu, H., Liu, H., Erasmus, S. W., Zhao, S., Wang, Q., & van Ruth, S. M. (2020). Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics. *Industrial Crops and Products, 158*. https://doi.org/10.1016/j.indcrop.2020.112956.

Yu, H., Liu, H., Wang, Q., & van Ruth, S. (2020). Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. *LWT - Food Science and Technology*, 109398. https://doi.org/10.1016/j.lwt.2020.109398.

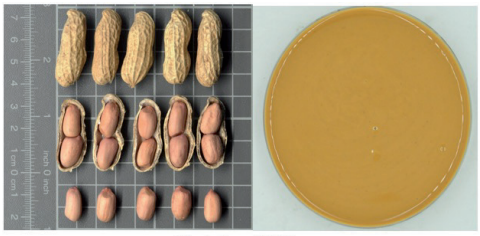
Zhang, S., Hu, X., Miao, H., Chu, Y., Cui, F., Yang, W., Wang, C., Shen, Y., Xu, T., Zhao, L., Zhang, J., & Chen, J. (2019). QTL identification for seed weight and size based on a high-density SLAF-seq genetic map in peanut (Arachis hypogaea L.). *BMC Plant Biology, 19*(1), 537. https://doi.org/10.1186/s12870-019-2164-5.

Zhao, J., Wang, T., Xie, J., Xiao, Q., Cheng, J., Chen, F., Wang, S., & Sun, B. (2019). Formation mechanism of aroma compounds in a glutathione-glucose reaction with fat or oxidized fat. *Food Chemistry, 270*, 436-444. https://doi.org/10.1016/j.foodchem.2018.07.106.

Zhao, Z., Shi, A., Wang, Q., & Zhou, J. (2019). High Oleic Acid Peanut Oil and Extra Virgin Olive Oil Supplementation Attenuate Metabolic Syndrome in Rats by Modulating the Gut Microbiota. *Nutrients, 11*(12). https://doi.org/10.3390/nu11123005.

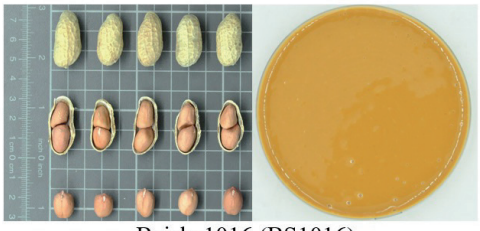
Zhou, D. N., Zhang, B., Chen, B., & Chen, H. Q. (2017). Effects of oligosaccharides on pasting, thermal and rheological properties of sweet potato starch. *Food Chemistry, 230*, 516-523. https://doi.org/10.1016/j.foodchem.2017.03.088.

SUPPLEMENTARY MATERIALS

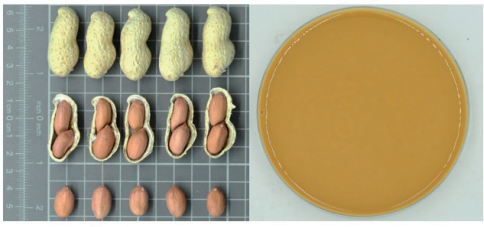


Baisha (BS)

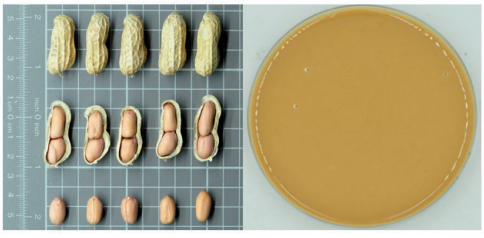
DF06 (high oleic acid variety)



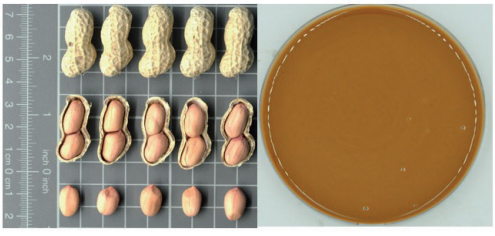
Baisha1016 (BS1016)



Fenghua1 (FH1)



Fuhua12 (FH12)



Huayu20 (HY20)

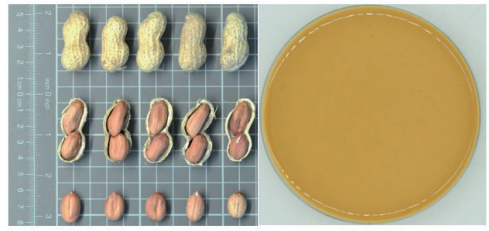
Hanghua2 (HH2)

Huayu22 (HY22)

Huayu25 (HY25)

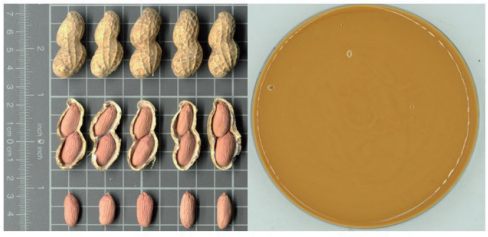
Huayu661 (HY661, high oleic acid variety)


Huayu33 (HY33)

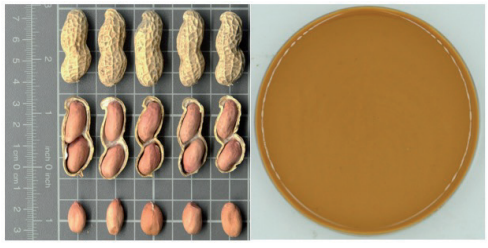

Huayu917 (HY917, high oleic acid variety)

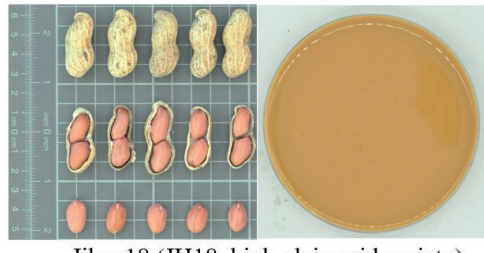

Huayu962 (HY962, high oleic acid variety)


Jihua5 (JH5)


Huayu963 (HY963, high oleic acid variety)


Jihua8 (JH8)


Jihua10 (JH10)


Jihua18 (JH18, high oleic acid variety)


Jihua16 (JH16, high oleic acid variety)


Jihua19 (JH19, high oleic acid variety)

Jihuatian1 (JHT1)

Jinhua9 (JinH9)

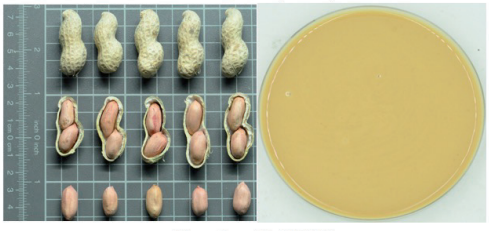
Jinhua8 (JinH8)

Luhua11 (LH11)

Puhua 28 (PH28)

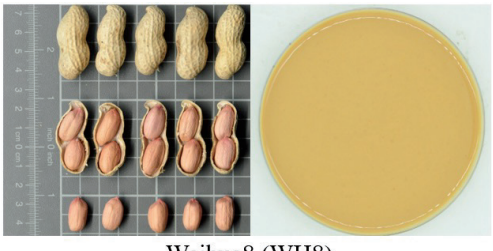
Shanhua8 (SH8)

Shanghua7 (SH7)



Shanhua9 (SH9)

Shanhua11 (SH11)



Weihua8 (WH8)

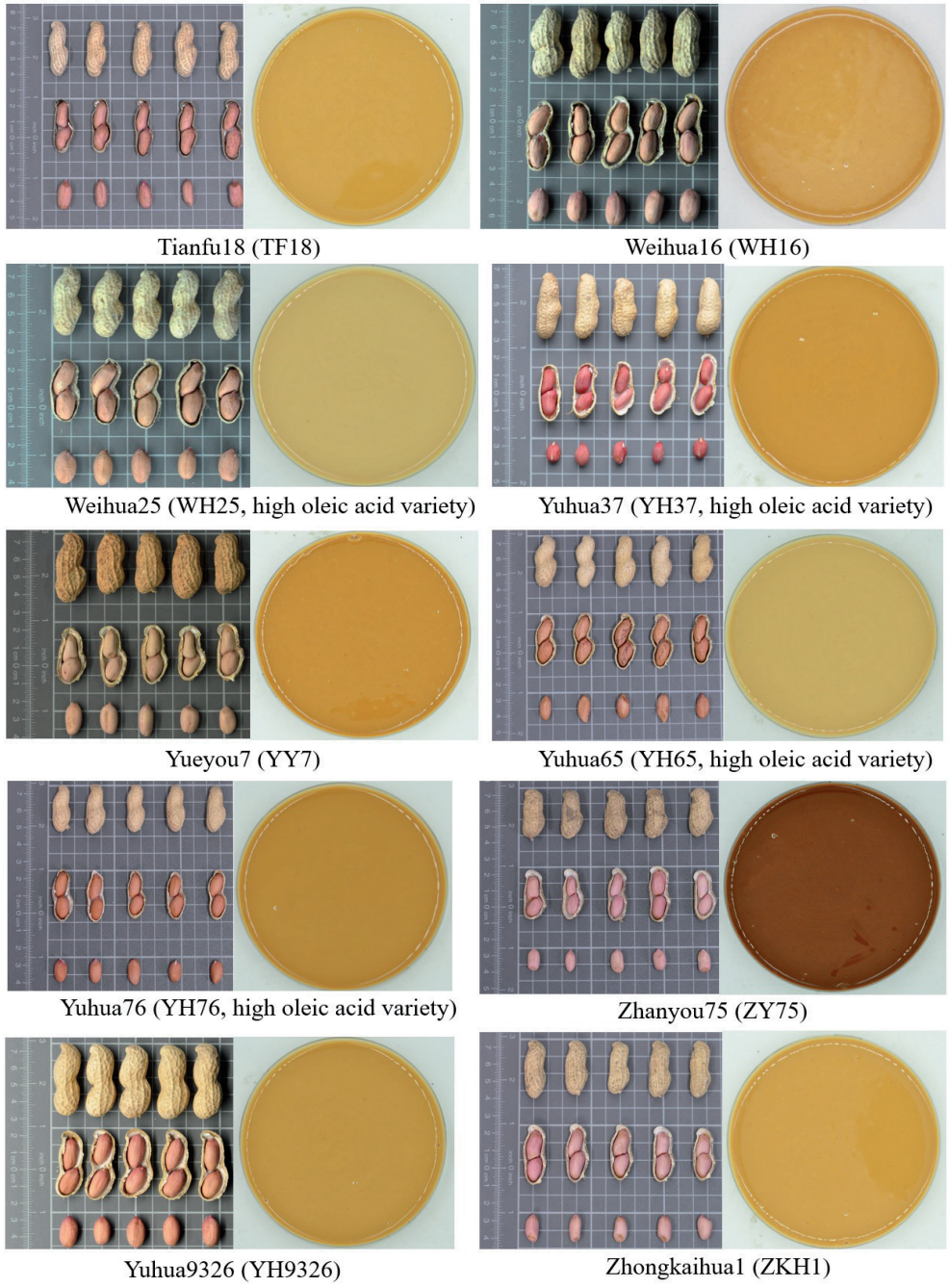
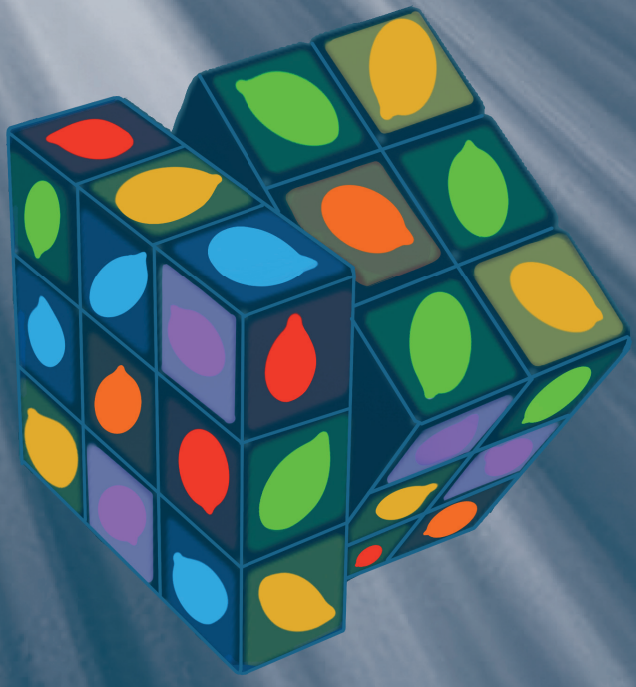


Figure S4.1. The pictures of peanuts and peanut butters (n=40)



CHAPTER 5

Rapid classification of peanut varieties for their processing into peanut butters based on near-infrared spectroscopy combined with machine learning

This chapter has been submitted to a journal as:
Yu, H., Liu, H., Erasmus, S. W., Wang, Q., & van Ruth, S. M. Rapid classification of peanut varieties for their processing into peanut butters based on near-infrared spectroscopy combined with machine learning.

ABSTRACT

Near-infrared spectroscopy (NIRS) for classifying peanut varieties for their processing into peanut butters was assessed for the first time. The principal component analysis (PCA) combined with cluster analysis based on the structure characteristics (texture and rheology) and roast characteristics (colour and volatile compounds) of the resulting peanut butters were conducted to group peanut varieties. After spectral collection of raw materials and different pre-treatments, partial least squares discrimination analysis (PLS-DA), support vector machine (SVM), and random forest (RF) as modelling methods were studied for comparison. PCA, variable importance (VarImp), and random forest selection by filter (RFSBF) as feature extraction methods were investigated. The sensitivity, specificity, and accuracy of the filtered cross validation and external validation models were all over 90%, while the kernel density estimation (KDE) analysis presented significant distributions of class probabilities between different groups. These results showed that NIRS combined with machine learning methods is a promising approach to provide a reliable evaluation of peanuts for precision processing.

Keywords: Classification; Near-infrared spectroscopy; Peanut butters; Random forest; Support vector machine.

5.1. INTRODUCTION

Peanuts are one of the paramount nuts in the world with the total production in 2019 equalling 48.8 million tons (Faostat, 2020). Peanuts can be consumed as either raw materials or as varieties processed products, such as roasted peanuts, peanut oil, and peanut butter, to satisfy consumer preferences and nutritional requirements (Wang, 2018). Among them, peanut butter with its characteristic mouthfeel and attractive flavour is a popular product all over the world (Gong et al., 2018). The industrial scale production of peanut butter is fully matured and commonly processed through roasting and grinding of peanuts (Wang, 2016). Peanuts impact the structure and roast characteristics of their resulting peanut butter given that different peanut varieties have different chemical compositions (Dhamsaniya et al., 2012). For precision processing, peanut varieties could be classified based on the processing suitability linked with the characteristics of peanut butter (Wang et al., 2017). Typically, the characteristics of peanut butter are evaluated based on its structure characteristics (texture and rheology) and roast characteristics (colour and volatile compounds) through time-consuming and costly laboratory analyses (Shakerardekani et al., 2013). Hence, the development of speedy, precise, and stable methods is vital to select suitable peanuts for processing to match the growing need for high-quality peanut butter production.

Currently, near-infrared spectroscopy (NIRS) based on the vibration of hydrogen-containing molecules has been systematically applied for the quality evaluation of agricultural products. On one hand, it has obvious advantages as it can rapidly collect and analyse spectral data without laborious preparations and it is also an environmentally and economically friendly method without chemical waste and high expenditure compared with the chemical methods. On the other hand, it initially needs stable calibration models built through representative samples collection, spectral and reference data analysis, and model establishment. Previously, abundant studies have been carried out to analyse the quality of peanuts, and satisfactory results were obtained for fat (Yu et al., 2016), protein and protein subunits (Zhao et al., 2021), amino acids (Wang et al., 2013), sucrose (Yu, Liu, Erasmus, et al., 2020), and fatty acids (Yu, Liu, Wang, et al., 2020). Meanwhile, NIRS was also used to grade peanuts (Sundaram et al., 2009) and determine their maturity (Windham et al., 2010). Evidently, NIRS can be used to evaluate the properties of peanuts, while the characteristics of peanut butter are related to the raw materials used. Therefore, it can be hypothesised that the spectral data of peanuts could be applied to reflect the characteristics of peanut butters at least under the decided preparation process. However, the challenge still exists to predict the characteristics of peanut butters through the acquired spectral data of their corresponding raw materials.

Hence, a suitable approach for this study was to firstly conduct a cluster analysis of peanut butters based on the characteristics analysis results to achieve scientific grouping of the corresponding peanut varieties. Secondly, fast classification models of peanuts were built combined with spectral data through chemometrics and machine learning. As a fact, machine learning such as random forest (RF) has been developed to obtain better classification and regression models because of its ability to deal with complex systems and multi-variables (Monforte et al., 2021; Phan & Tomasino, 2021). Therefore, machine learning has the huge potential to analyse the spectral data of peanuts and their relationship with peanut butters.

To build better performances classification models, RF, support vector machine (SVM), and partial least squares discriminant analysis (PLS-DA) as modelling algorithms were compared in this study based on the different pre-treatment spectral data. Principal component analysis (PCA), variable importance (VarImp) (Kuhn, 2012), and random forest selection by filter (RFSBF) (Kuhn & Johnson, 2013) were also conducted and compared for feature extraction to establish the simplified and stable models. The sensitivity, specificity, and accuracy were used to assess the model performances, while a kernel density estimation (KDE) assessed the distributions of class probabilities.

5.2. MATERIALS AND METHODS

5.2.1. Peanuts and peanut butters

A total of 40 peanut varieties were listed in **Table S5.1**, which includes the main-planting varieties and high oleic acid varieties. For each variety, 5 kg samples were collected and stored at commercial 4°C cold storage (Yuandong Co., Ltd., Tianjin, China). The peanut butters were prepared according to a general process (Yu et al., 2021). Concisely, approximately 0.5 kg plump peanut kernels per variety were placed on a steel bakeware and roasted at 160°C for 30 min in an electric oven with the top and bottom heating mode. The peeled roasted peanuts were then ground in a colloid grinder, producing about 0.4 kg peanut butter per variety for analysis which was stored at room temperature (21°C).

5.2.2. Spectral data collection

A benchtop spectrometer with a rotating sphere (Bruker Scientific Instruments, Karlsruhe, Germany) was used to acquire the spectral data. About 100 g kernels were placed in a ring cup (9.7 cm diameter and 4.5 cm depth) for each sample measurement of which each spectrum was the average of 32 scans. Each variety was sampled five times, producing a total of 200 (40 × 5) spectra for the subsequent analysis. The reflectance spectra

were recorded by an indium gallium arsenide (InGaAs) detector where the wavelength ranged from 12489.49 cm^{-1} to 3996.02 cm^{-1}. Spectral acquisition and conversion were conducted using OPUS 7.5 software (Bruker Scientific Instruments, Karlsruhe, Germany).

5.2.3. Physicochemical characteristics of peanut butters

5.2.3.1. Texture

Peanut butter texture was described using firmness (g). A TA-XTplus texture analyser (Micro Stable System Co., UK) with the back-extrusion penetration model was used to measure the texture characteristics. The analyser was coupled with a long roller and a 35 mm diameter compression disc. Each sample was put into a cylindrical jar (50 mm internal diameter) with the same volume (75%). The trigger pressure of penetration was 5 g, and the penetration depth was 30 mm with a speed of 1 mm/s. Samples were analysed in duplicate to measure the force as firmness and then averaged to obtain the mean value per sample.

5.2.3.2. Rheology

An HR-2 rheometer (TA Instruments, New Castle, the USA) was applied to evaluate the rheological characteristics of the peanut butters. Approximately 2 g sample was positioned on the crosshatched plate (40 mm diameter) with 1 mm gap geometry and subsequently assessed. After 2 min of equilibration, steady-state detection was conducted at a shear rate range (1-300/s) for the shear rate sweep test. The Herschel-Bulkley's model was applied to model the flow behaviour of the peanut butters (Ahmed & Ramaswamy, 2006). The yield stress as an important parameter of the model means the lowest shear stress required to trigger peanut butter to flow, which was confirmed for the next analysis. The storage modulus (*G'*) and loss modulus (*G"*) were confirmed by dynamic oscillatory experiments. All experiments were performed inside the linear viscoelastic range determined at the 1 Hz frequency. The results were collected in the frequency range of 0.1 to 100 Hz. As indicated in Eq. (1) and Eq. (2), the *G'* and *G"* were fitted by power function law equations of oscillatory frequency (ω), expressing the viscoelastic characteristics of peanut butter.

$$G' = a \times \omega^b \quad (1)$$

$$G'' = c \times \omega^d \quad (2)$$

Where a (kPa Hz^{-b}) and c (kPa Hz^{-d}) demonstrate the quantity of G' and G" correspondingly at a specific frequency, and b and d (×100) indicate the slope of the connections between the modulus and frequency (Liu et al., 2019; Resch & Daubert, 2002). Samples were analysed in duplicate and then averaged to obtain the mean value per sample.

5.2.3.3. Colour

Peanut butter colour was confirmed using a CS-600 portable colour spectrophotometer (CHNSpec Technology Co., Ltd, Hangzhou, China). For each colour assessment, a 2 g sample of each peanut butter was put in a circular quartz cell and subsequently the L* (darkness at 0 to lightness at 100), a* (greenness at -128 to redness at 127), and b* (blueness at -128 to yellowness at 127) colour values were determined. Triplicate measurements were performed and the mean value was calculated for each peanut butter sample.

5.2.3.4. Volatile compounds

The volatile compounds measurement of the peanut butter samples was based on headspace solid-phase microextraction-gas chromatography-mass spectrometry (HS-SPME-GS-MS). The whole analysis protocol could be found in our previous study (Yu et al., 2021). Briefly, samples were prepared by weighing 5 g of each peanut butter into a 20 mL glass vial, and 50 µL internal standard 1,2,3-trichloropropane (0.5 mg/mL in methanol, Sinopharm Chemical Reagent Co., Ltd., Beijing, China) was added to each sample vial for the concentration calculation. Each vial was sealed with a Teflon diaphragm and an aluminium lid. The samples were put in shaking incubators at 55°C for 10 min pre-equilibrium after which the SPME fibre was introduced to the headspace for 40 min. The absorbed volatiles were transferred in the hot injection section (260°C) for 150 s desorption. The splitless mode was applied for the GS-MS analysis with helium as the carrier gas and a flow rate of 1 mL/min. The analyser's temperature was set at 250°C, while the oven temperature program was initiated at 40°C for 5 min and subsequently raised at a rate of 5 °C/min to 250°C with a holding time of 5 min. Mass spectra were obtained using the electron impact ionisation mode (70 eV) in the mass range of 35 to 500 m/z. By comparing the data to the mass spectral library and calculating the retention indices (RI) (Yu et al., 2021), volatile compounds were recognized. The calculation of RI, as shown in the Eq. (3), relies on the n-alkane standard (C7-C40) (0.5 mg/mL in n-hexane, Smart Solutions, North Charleston, American) as the reference.

$$\mathrm{RI_x} = 100\mathrm{n} + 100 \times {(\mathrm{tR_x} - \mathrm{tR_n})}\Big/{(\mathrm{tR_{n+1}} - \mathrm{tR_n})} \tag{3}$$

Where tR indicates the retention time, n indicates the number of atom carbon, and x indicates one of volatile compounds. After that, the effective volatile compounds were confirmed by calculating the ratio of the concentrations / the odour threshold (>1) (Yu et al., 2021). Duplicate measurements were performed and the mean value was calculated for each peanut butter sample.

5.2.4. Chemometrics and machine learning

After all the physicochemical characteristics analyses were performed on the peanut butter samples, principal component analysis (PCA) was conducted to reduce data dimension to explore the relationships of the peanut butter characteristics. The outliers of peanut varieties were estimated by the Hotelling's T^2 based on the PCA results (Liu et al., 2018). K-means as unsupervised clustering was applied to acquire the distinct groups as the reference values. Kruskal-Wallis tests were used to assess whether the structure and roast characteristics of the various groups differed significantly. The strength of the connections between the characteristics of peanut butters and the absorbances of the original spectral data was assessed using correlation analysis.

All spectral datasets were prior pre-processed in various methods to improve the signal-to-noise ratio and uncover more relevant data, including standard normal variate (SNV), the first derivative (FD), the second derivative (SD), normalisation, and multiple scatter correlation (MSC). The spectral datasets were split using a 4:1 ratio into a training dataset and a validation dataset. The three repeats of 10-fold cross validation were used to prevent model overfitting. Partial least squares discrimination analysis (PLS-DA), support vector machine (SVM), and random forest (RF) were used for mathematical modelling. Since there are 1102 spectral variables, some features were extracted by PCA, variable importance (VarImp), and random forest selection by filter (RFSBF) to simplify the models. PCA converts the original spectral data into new orthogonal and non-overlapping principal components (PCs). The retrieved features in this research were the sum of the top five PCs, accounting for 99% of total variances. VarImp scores were generated to determine the feature importance by using the caret package (Kuhn, 2008). The variables (score value >1) were selected as the important spectra. RFSBF in the caret package with 10-fold cross validation tests were applied to select the best feature spectra. The performances of all classification models were assessed based on the following several indicators: accuracy (ACCU, the total rate of accurate classification); sensitivity (SENS, the rate of accurate confirmation); and specificity (SPEC, the rate of accurate rejection). A kernel density estimation function (KDE) was used to generate the genuinely positive and negative rate distribution. KDE is a kind of non-parametric distribution assessment that is similar to histogram, but allows for distribution interpolation and modification (Alewijn et al., 2016). All calculations were performed by R 4.0.3 software (R Foundation for Statistical Computing, Austria).

5.3. RESULTS AND DISCUSSION

5.3.1. The cluster results of peanut butters and the spectral results of peanut varieties

Figure 5.1a shows the cluster analysis results of peanut butters based on the textural and rheological characteristics. The first two principal components (PCs) accounted for 90% of the total variance and therefore contained most of the relevant information. Group 1 (blue) was located in the positive direction of PC1, indicating that it had positive relationships with firmness, field stress, *G*'-a, and *G*"-c, while group 2 (red) was located in the negative direction of PC1. Correspondingly, group 1 had significantly higher ($P < 0.05$) values for firmness (43.4 ± 8.0 g), yield stress (11.13 ± 1.81 Pa), *G*'-a (60.94 ± 14.01 kPa Hz^{-b}), and *G*"-c (48.59 ± 6.99 kPa Hz^{-d}) than group 2 (34.6 ± 5.6 g, 6.12 ±1.83 Pa, 32.61 ± 9.94 kPa Hz^{-b}, and 24.68 ± 6.34 kPa Hz^{-d}, respectively). Firmness is related to the hardness of peanut butters, while field stress, *G*'-a, and *G*"-c present the flow and deformation of peanut butters under stress and strain (Sun & Gunasekaran, 2009). The textural and rheological qualities reflect the structural state of peanut butter which is connected to the sensory attributes evaluated (Shakerardekani et al., 2013).The clustering results clearly show that there were two groups of peanut butters based on the texture and rheological characteristics.

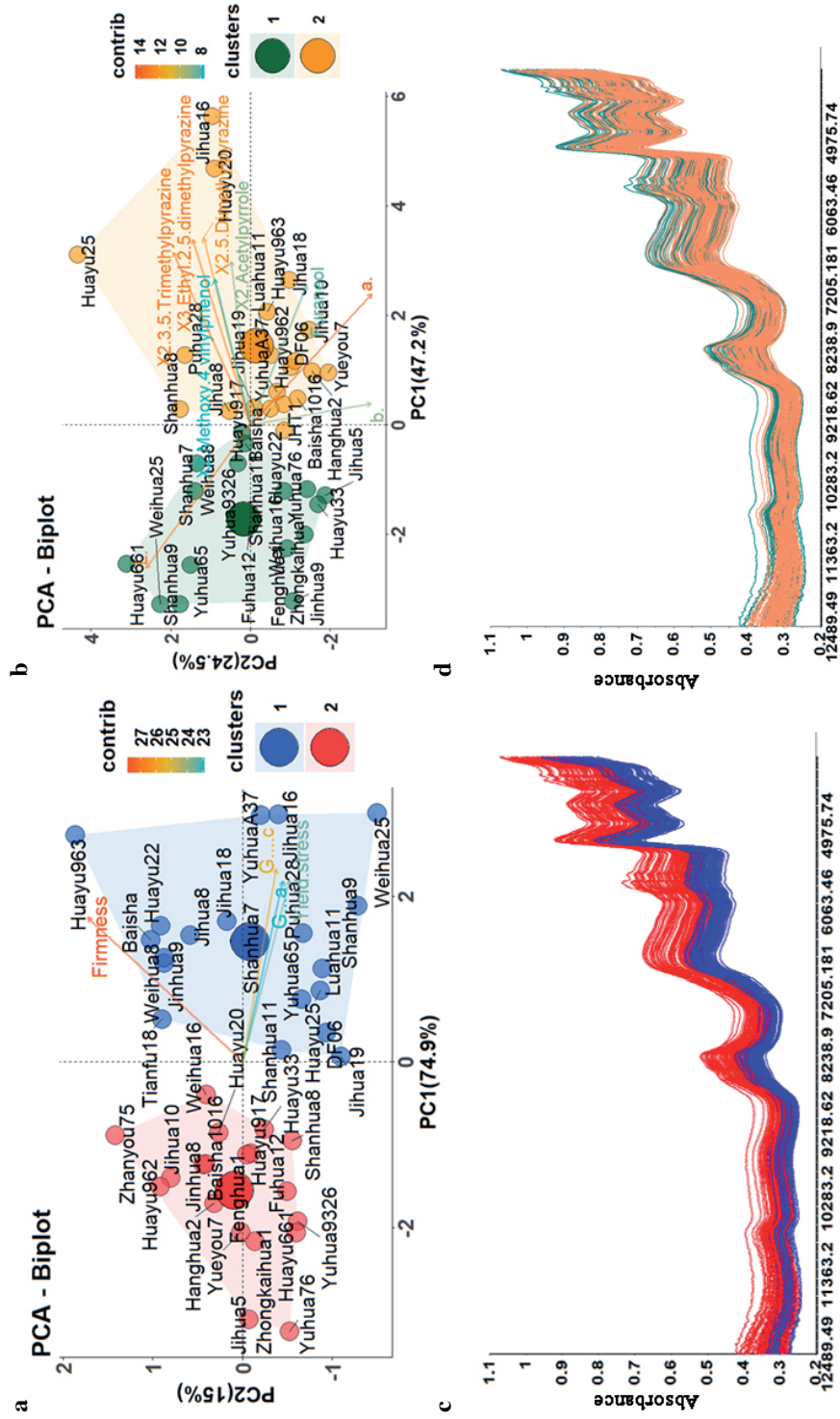
Based on the cluster results, the spectral data of the raw materials are shown in **Figure 5.1c**. Similarly, the blue lines stood for group 1, while the red lines stood for group 2. Overall, the spectral absorbance values of group 2 were higher than that of group 1. The spectral absorbance values derived from the C-H, O-H group of fat, protein, and sucrose in peanuts (Hourant et al., 2000) which is known to have a great influence on the texture and rheological characteristics of peanut butters (Dhamsaniya et al., 2012; Mohd Rozalli et al., 2015). PCA was used to explore the spectral data of peanuts to reduce multicollinearity. The sum of PC1 and PC2 was 92%. Group 1 (blue) was mostly grouped in the right direction of PC1, while group 2 (red) was located on the left direction in **Figure 5.1e**, which had the same trend as PCA results in **Figure 5.1a**. Therefore, it shows the feasibility to classify peanut butters based on the spectra data of peanut varieties.

Figure 5.1g presents the strength of the relationship between the characteristics of peanut butter (vertical) and the full wavelength (horizontal). Compared with the results in **Figure 5.1a**, yield stress, *G*'-a, and *G*'-c had a similar relationship with the wavelength. Specifically, these three indicators had higher negative correlation coefficient values with longer wavelength ranges (8786.62 cm^{-1}– 3996.02 cm^{-1}) and positive correlation coefficient values with short wavelength ranges (12489.49 cm^{-1} – 11100.92 cm^{-1}). Firmness had positive correlation values with short wavelength ranges (12489.49 cm^{-1} – 11100.92 cm^{-1}). Group 1 had higher these indicators, resulting in lower absorbance in the longer wavelength ranges shown in **Figure 5.1c**. Wavelengths in these ranges are likely derived from

the 1st overtone region, the 3rd overtone region, and the combination bands of the main chemical compositions (fat, protein, and sucrose) in peanuts (Barbin et al., 2014).

The cluster results of peanut varieties based on the roast characteristics including colour and volatile compounds are presented in **Figure 5.1b**. PC1 and PC2 accounted for 72% of the total variance. Group 1 (green) was located in the negative direction of PC1, where L* made the great contributions, while group 2 (yellow) was located in the positive direction along PC1 where the major contributions were derived from most of the roast characteristics such as pyrazines, 2-acetylpyrrole, and a*. There were also significant differences ($P < 0.05$) in the roast characteristics between group 1 and group 2. Specifically, group 2 had higher 2,5-dimethylpyrazine (6.19 ± 1.97 mg/kg), 2,3,5-trimethylpyrazine (2.65 ± 1.01 mg/kg), 3-ethyl-2-5-dimethylpyrazine (1.79 ± 0.68 mg/kg), furaneol (1.68 ± 0.72 mg/kg), 2-acetylpyrrole (1.15 ± 0.30 mg/kg), 2-methoxy-4-vinylphenol (2.54 ± 0.82 mg/kg) than group 1 (3.90 ± 1.30 mg/kg, 1.66 ± 0.71 mg/kg, 1.00 ± 0.36 mg/kg, 0.86 ± 0.49 mg/kg, 0.68 ± 0.31 mg/kg, and 1.53 ± 0.45 mg/kg, respectively). These volatile compounds, especially pyrazine with nutty and roast aromas, are known as some of the primary volatiles of peanut products (Baker et al., 2003; Li & Hou, 2018). The L* value (51.40 ± 4.68) of group 1 was lower than that of group 2 (57.04 ± 3.39 and 8.16 ± 2.06), while the a* value (10.94 ± 2.25) of group 1 was higher than that of group 2 (8.16 ± 2.06). Therefore, the manufacturers can select peanut varieties from group 2 to process peanut butters with rich flavour and bright colour.

The spectra of different groups based on the roast characteristics cluster analysis can be seen in **Figure 5.1d**. Most of the green lines (group 1) had higher absorbance values than the yellow lines (group 2), but they were not completed separated compared with the cluster results of the structural characteristics. The PCA results (**Figure 5.1f**) show more distinct relationships between the two groups. Although the total of PC1 and PC2 was 98% and group 2 was mainly located on the left side along PC1, there was no clear separation between the two groups. The main reason for the incomplete separation is because the roast characteristics of peanut butters are decided by raw materials and the roasting process, while the spectral data of raw materials only show partial information. Therefore, it cannot fully reflect the roast characteristics of peanut butters. The correlation heatmap offered more detailed information about the relationship between the spectral data of peanut varieties and the roast characteristics of the resulting peanut butters. It was found that a* value and 2-methoxy-4-vinylphenol had a significantly positive correlation ($P < 0.05$) with the whole wavelength. This is because a* value stands for the redness mechanically linked with the infrared band where the NIRS wavelength exists. There were also significant negative correlations ($P < 0.05$) between 2,5-dimethylpyrazine and 3-ethyl-2-5-dimethylpyrazine and the longer wavelength ranges (11332.34 cm^{-1} - 3996.02 cm^{-1}). This could imply that the spectral data of raw materials can reflect the characteristics of peanut butters.


a
PCA - Biplot
PC1(74.9%)
PC2(15%)
contrib
clusters
b
PCA - Biplot
PC1(47.2%)
PC2(24.5%)
contrib
clusters
c
Absorbance
12489.49 11363.2 10283.2 9218.62 8238.9 7205.181 6063.46 4975.74
d
Absorbance
12489.49 11363.2 10283.2 9218.62 8238.9 7205.181 6063.46 4975.74

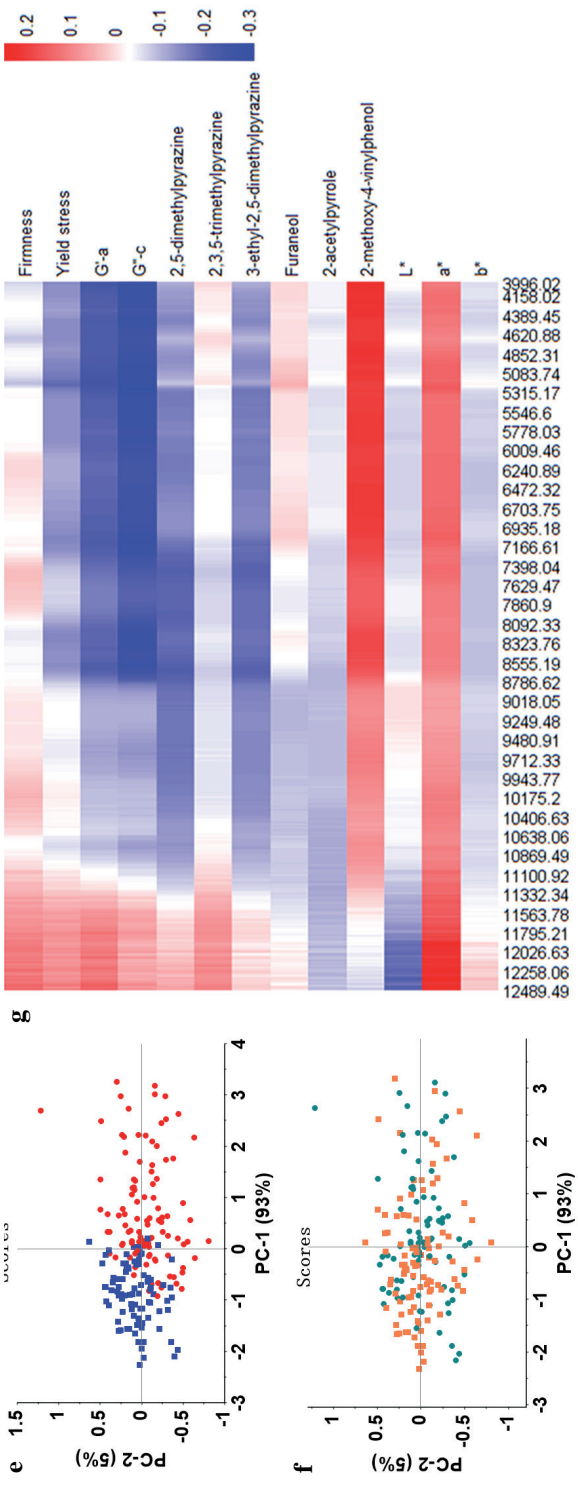


Figure 5.1. The principal component analysis (PCA) biplot of the structure characteristics (a) and the roast characteristics (b) of peanut butters. Different colored polygons indicate the different sample clusters and arrows indicate the contributions (contrib) of parameters. The raw spectral data of peanut varieties for structure characteristics (c) and roast characteristics (d) (vertical axis is absorbance and horizontal axis is the wavelength range from 12489.49 cm^{-1} to 3996.02 cm^{-1}). The PCA score plot of the spectral data of peanut varieties for structure characteristics (e) and roast characteristics (f). The (g) correlation analysis between raw spectral data of peanut varieties (vertical axis) and characteristics of peanut butters (horizontal axis). Positive and negative coefficients are coloured red and blue, respectively. coefficients < -0.15 and > + 0.15 indicate significant correlations ($P < 0.05$).

5.3.2. The classification models built based on the full wavelength range

The PLS-DA, SVM, and RF were used to establish classification models based on different pre-treatment spectral datasets. The results are shown in **Table 5.1**. Overall, all training and cross validation models had good results. The SENS, SPEC, and ACCU of training models were over 97%, while the results of cross validation models were over 90% except for a few models. The performances of external validation models were quite unequal that the range of SENS, SPEC, and ACCU were 65% - 85%, 50% - 100%, and 60% - 93%, respectively. Therefore, external validation models were the key models to determine which algorithms have the best performances.

In terms of the modelling algorithms, PLS-DA is an effective, multivariate regression-based algorithm for peanuts classification. Although compelling, PLS-DA incurs performance degeneration under complex situations such as class imbalance and multiclass, which are common in peanut varieties (Song et al., 2018). The full wavelength including 1102 variables, has the nonlinearity effects on the PLS-DA models. Therefore, SVM and RF as non-parametric machine learning algorithms have more advantages. Although the cross validation results showed that SVM had better results than RF, the external validation results had lower performance than RF, which meant that SVM models existed the overfitting effects. Hence, RF can avoid overfitting and had the best prediction results compared to other algorithms, while the average SENS, SPEC, and ACCU of RF were 77%, 100%, and 89%, respectively.

Concerning the pre-treatment methods, all methods increased the performances of the models. Among them, FD had the best improvements in the performances of models. The average performances (SENC, SPEC, and ACCU) of models based on FD were 93%, 94%, and 93% (respectively) for cross validation, and 77%, 95%, and 86% (respectively) for external validation. The results showed that the model established based on the full wavelength has great performances to classify peanut varieties to produce different structural characteristics of peanut butters.

The classification models for classifying peanuts varieties based on the roast characteristics (colour and volatile compounds) are compared in **Table 5.2**. The performances of classification models for roast characteristics were overall not good when compared with the classification models for structure characteristics. This is because the roast characteristics are not only derived from the raw materials but are also generated by brown reactions and caramelisation during the roasting process. Therefore, the spectral data of the raw materials lacked the information about roasting, leading to less control over controlling roasting variation. Among all pre-treatment methods, SNV was the best pre-treatment and the average performances (SENC, SPEC, and ACCU) were 88%, 85%, and 86%

Table 5.1. The sensitivity, specificity, and accuracy results for the discrimination of peanut varieties based on the structure characteristics of peanut butters combined with the different spectral pre-treatment and algorithm methods.

Algorithms	Pre-treatments	Training (n=155)			Cross validation			External validation (n=40)		
		SENS(%)	SPEC(%)	ACCU(%)	SENS(%)	SPEC(%)	ACCU(%)	SENS(%)	SPEC(%)	ACCU(%)
PLS-DA	Original	100%	98%	99%	87%	89%	88%	70%	50%	60%
	SNV	100%	100%	100%	92%	93%	93%	75%	65%	70%
	FD	100%	100%	100%	89%	94%	91%	80%	90%	85%
	SD	100%	100%	100%	83%	86%	84%	70%	75%	73%
	MSC	98%	100%	99%	95%	89%	92%	70%	95%	83%
	Normalisation	99%	100%	99%	95%	93%	94%	70%	85%	78%
SVM	Original	98%	97%	97%	93%	91%	92%	80%	100%	90%
	SNV	100%	100%	100%	99%	99%	99%	65%	100%	83%
	FD	100%	100%	100%	96%	95%	96%	70%	95%	83%
	SD	100%	100%	100%	94%	97%	95%	65%	95%	80%
	MSC	100%	100%	100%	98%	99%	98%	65%	100%	83%
	Normalisation	98%	97%	97%	95%	92%	94%	75%	100%	88%
RF	Original	100%	100%	100%	89%	90%	90%	80%	100%	90%
	SNV	100%	100%	100%	95%	97%	96%	65%	100%	83%
	FD	100%	100%	100%	93%	93%	93%	80%	100%	90%
	SD	100%	100%	100%	89%	94%	91%	65%	100%	83%
	MSC	100%	100%	100%	94%	96%	95%	85%	100%	93%
	Normalisation	100%	100%	100%	91%	91%	91%	85%	100%	93%

(ACCU) Accuracy; (FD) The first derivative; (MSC) Multiple scatter correlation; (n) Number of samples; (PLS-DA) Partial least squares discrimination analysis; (RF) Random forest; (SD) The second derivative; (SENS) Sensitivity; (SPEC) Specificity; (SNV) Standard normal variate; (SVM) Support vector machine.

Table 5.2. The sensitivity, specificity, and accuracy results for the discrimination of peanut varieties based on the roast characteristics of peanut butters combined with the different spectral pre-treatment and algorithm methods.

		Training (n=148)			Cross validation			External validation (n=37)		
Algorithms	**Pre-treatments**	**SENS(%)**	**SPEC(%)**	**ACCU(%)**	**SENS(%)**	**SPEC(%)**	**ACCU(%)**	**SENS(%)**	**SPEC(%)**	**ACCU(%)**
PLS-DA	Original	91%	96%	94%	75%	70%	71%	80%	88%	84%
	SNV	99%	94%	97%	76%	75%	75%	85%	94%	89%
	FD	81%	79%	80%	66%	63%	65%	50%	76%	62%
	SD	100%	100%	100%	66%	64%	65%	55%	82%	68%
	MSC	94%	88%	91%	79%	72%	75%	75%	88%	81%
	Normalisation	94%	91%	93%	72%	69%	71%	85%	88%	86%
SVM	Original	100%	100%	100%	84%	81%	82%	65%	76%	70%
	SNV	100%	100%	100%	98%	97%	97%	100%	94%	98%
	FD	100%	100%	100%	80%	76%	78%	100%	71%	86%
	SD	100%	100%	100%	75%	73%	74%	65%	59%	62%
	MSC	100%	100%	100%	96%	99%	98%	85%	65%	76%
	Normalisation	100%	100%	100%	82%	76%	80%	65%	77%	70%
RF	Original	100%	100%	100%	66%	57%	62%	70%	41%	57%
	SNV	100%	100%	100%	90%	84%	87%	75%	88%	81%
	FD	100%	100%	100%	76%	75%	75%	95%	82%	89%
	SD	100%	100%	100%	76%	53%	65%	65%	47%	57%
	MSC	100%	100%	100%	90%	85%	88%	60%	88%	73%
	Normalisation	100%	100%	100%	66%	63%	64%	65%	35%	51%

(ACCU) Accuracy; (FD) The first derivative; (MSC) Multiple scatter correlation; (n) Number of samples; (PLS-DA) Partial least squares discrimination analysis; (RF) Random forest; (SD) The second derivative; (SENS) Sensitivity; (SPEC) Specificity; (SNV) Standard normal variate; (SVM) Support vector machine

(respectively) for cross validation, and 87%, 92%, and 89% (respectively) for external validation. In respect of the modelling algorithms, SVM and RF showed better performances when compared to the PLS-DA algorithm. Among all models, SNV-SVM had the best performances (SENC, SPEC, and ACCU) with all parameters 100% for the training models, 98%, 97%, and 97% (respectively) for the cross validation models, and 100%, 94%, and 98% (respectively) for external validation models.

5.3.3. The classification models built based on the features extracted

The results of PLS-DA, SVM, and RF models based on the features extracted by PCA, VarImp, and RFSBF are shown in **Tables 5.3** and **5.4**. PCA is the conventional method to reduce the dimension of data thought building new non-linear variables. The classification models for the structure characteristics are shown in **Table 5.3**. The top five PCs of the FD spectral data were selected as the new variables to build models. The PCA-RF models had the best prediction performances compared to the other methods. The ACCU of training, cross validation, and external validation models were 100%, 95%, and 88%, respectively, which was similar to the results obtained for the full wavelength models. For VarImp and RFSBF, 19 wavelengths and 719 wavelengths were selected as features, respectively. The VarImp method estimated model performances by using the minimum number of wavelengths. The ACCU of external validation models were 90%, 80%, and 83% for PLS-DA, SVM, and RF, respectively. RFSBF method maintained or improved the performances of the full wavelength models. Especially, FD-RFSBF-SVM and FD-RFSBF-RF models had better performances. The ACCU of training, cross validation, and external validation models were 100%, 96%, and 90% for FD-RFSBF-SVM and 100%, 93%, and 88% for FD-RFSBF-RF. There are slight differences among different models. Relatively speaking, FD-RFSBF-SVM had the best performances. These results proved that features extracted by PCA, VarImp, and RFSBF could be effectively used to build high accuracy and belief models for classifying peanut varieties to process different structure characteristics of peanut butters.

The classification models of the roast characteristics based on the features are presented in **Table 5.4**. The five new variables extracted from SNV spectral data by the PCA methods were used to establish the models. The SNV-PCA-SVM had the best prediction performances compared to the other methods. The ACCU of training, cross validation, and external validation models were 99%, 96%, and 89%, respectively, which were in line with the full wavelength models. The performances of models built by the 728 feature wavelengths based on RFSBF were consistent with the results based on the corresponding whole wavelength models. The SNV-RFSBF-SVM models had the best performances with 100%, 97%, and 95% for the ACCU of training, cross validation, and external validation models, respectively. Only five wavelengths (4528.31 cm^{-1}, 4536.02 cm^{-1}, 10044.05 cm^{-1},

Table 5.3. The sensitivity, specificity, and accuracy results for the discrimination of peanut varieties based on the structure characteristics of peanut butters combined with the different algorithm methods using features extracted by PCA, RFSBF, and VarImp.

			Training (n=155)			Cross validation			External validation (n=40)		
Algorithms	**Data type**	**n***	**SENS(%)**	**SPEC(%)**	**ACCU(%)**	**SENS(%)**	**SPEC(%)**	**ACCU(%)**	**SENS(%)**	**SPEC(%)**	**ACCU(%)**
PLS-DA	Full	1102	100%	100%	100%	89%	94%	91%	80%	90%	85%
	PCA	5	88%	97%	93%	87%	95%	91%	80%	100%	90%
	RFSBF	719	94%	100%	97%	91%	92%	91%	80%	95%	88%
	VarImp	19	89%	96%	93%	89%	96%	92%	80%	100%	90%
SVM	Full	1102	100%	100%	100%	96%	95%	96%	70%	95%	83%
	PCA	5	99%	99%	99%	97%	98%	97%	75%	100%	88%
	RFSBF	719	100%	100%	100%	96%	96%	96%	80%	100%	90%
	VarImp	19	98%	100%	99%	94%	94%	94%	65%	95%	80%
RF	Full	1102	100%	100%	100%	93%	93%	93%	80%	100%	90%
	PCA	5	100%	100%	100%	93%	96%	95%	80%	95%	88%
	RFSBF	719	100%	100%	100%	93%	93%	93%	75%	100%	88%
	VarImp	19	100%	100%	100%	93%	93%	93%	65%	100%	83%

*n is the number of variables.

(ACCU) Accuracy; (n) Number of samples; (PCA) Principal component analysis; (PLS-DA) Partial least squares discrimination analysis; (RF) Random forest; (RFSBF) Random forest selection by filter; (SENS) Sensitivity; (SPEC) Specificity; (SNV) Standard normal variate; (SVM) Support vector machine; (VarImp) Variable importance.

Table 5.4. The sensitivity, specificity, and accuracy results for the discrimination of peanut varieties based on the roast characteristics of peanut butters combined with the different algorithm methods using features extracted by PCA, RFSBF, and VarImp.

			Training (n=148)			Cross validation			External validation (n=37)		
Algorithms	**Data type**	**n***	**SENS(%)**	**SPEC(%)**	**ACCU(%)**	**SENS(%)**	**SPEC(%)**	**ACCU(%)**	**SENS(%)**	**SPEC(%)**	**ACCU(%)**
PLS-DA	Full	1102	99%	94%	97%	76%	75%	75%	85%	94%	89%
	PCA	5	69%	56%	63%	69%	53%	60%	60%	53%	57%
	RFSBF	728	91%	94%	88%	80%	74%	77%	85%	94%	89%
	VarImp	5	80%	66%	74%	84%	62%	70%	55%	59%	57%
SVM	Full	1102	100%	100%	100%	98%	97%	97%	100%	94%	98%
	PCA	5	99%	99%	99%	95%	97%	96%	90%	88%	89%
	RFSBF	728	100%	100%	100%	98%	97%	97%	95%	94%	95%
	VarImp	5	96%	99%	97%	89%	85%	87%	75%	88%	81%
RF	Full	1102	100%	100%	100%	90%	84%	87%	75%	88%	81%
	PCA	5	100%	100%	100%	89%	85%	86%	85%	77%	81%
	RFSBF	728	100%	100%	100%	90%	82%	86%	75%	88%	81%
	VarImp	5	100%	100%	100%	86%	84%	84%	70%	77%	73%

*n is the number of variables.

(ACCU) Accuracy; (n) Number of samples; (PCA) Principal component analysis; (PLS-DA) Partial least squares discrimination analysis; (RF) Random forest; (RFSBF) Random forest selection by filter; (SENS) Sensitivity; (SPEC) Specificity; (SNV) Standard normal variate; (SVM) Support vector machine; (VarImp) Variable importance.

10051.77 cm^{-1}, and 10537.77 cm^{-1}) were selected to build models by VarImp. Although the performances of the VarImp models were relatively poor when compared to the other methods, VarImp greatly simplified the complexity of the models. It can thus be suggested that different feature extraction methods could be used to simplify models and SNV-RFSBF-SVM was the best method for peanut varieties classification for processing different roast characteristics of peanut butters.

5.3.4. The KDE distribution of the selected models

Figure 5.2 presents the KDE distribution of the selected models including FD-RF and FD-RFSBF-SVM for structure characteristics, and SNV-SVM and SNV- RFSBF-SVM for roast characteristics. In contrast with binary analysis, KDE distribution analysis offers more information than a single value. The conventional binary analysis classifies samples based on a threshold value. Peanuts with probability values below the threshold are grouped into one class, while peanuts with probability values above the threshold value are grouped into the other groups. Generally, the number of peanut varieties classified accurately will be demonstrated. However, KDE distribution plots further show the distance between probability values to the threshold. The smaller the distance, the higher the risk of misclassification (Yan et al., 2019).

In this study, the default threshold was 0.5. **Figure 5.2a** shows two sub-groups in all FD-RF models. Specifically, the main body of the two sub-groups is separately located on both sides and the location of the peak are far from the threshold value of 0.5 in the training model. KDE distribution for cross validation and external validation models were similar. Although there was some superposition between the two groups, the main body of different groups were separate. The KDE results of FD-RFSBF-SVM shown in **Figure 5.2b** had the better distribution. Two sub-groups in the training model were completely distributed at both ends. Regarding the cross validation and external validation, the tails at both sides were light and the location of the peak was far from the threshold value of 0.5. These results agreed with the results shown in **Table 5.3.**

In respect of the classification models for the roast characteristics, **Figure 5.2c** and **Figure 5.2d** show that the training models of SNV-SVM and SNV- RFSBF-SVM had the greatest distributions of class probability as the different groups located on both ends of the X-axis. For cross validation, group 1 and group 2 were mainly separated with a limit of threshold value 0.5 and the principal part of class probability were 0.05 - 0.25 for group 1 and 0.75 - 1.00 for group 2, respectively, which matched the model performances shown in **Table 5.2** and **Table 5.4**. One could also see that the external validation model of SNV-SVM had a great segregation between different groups with very little overlap, while the same model of SNV- RFSBF-SVM had a relatively poor distribution that each group had a

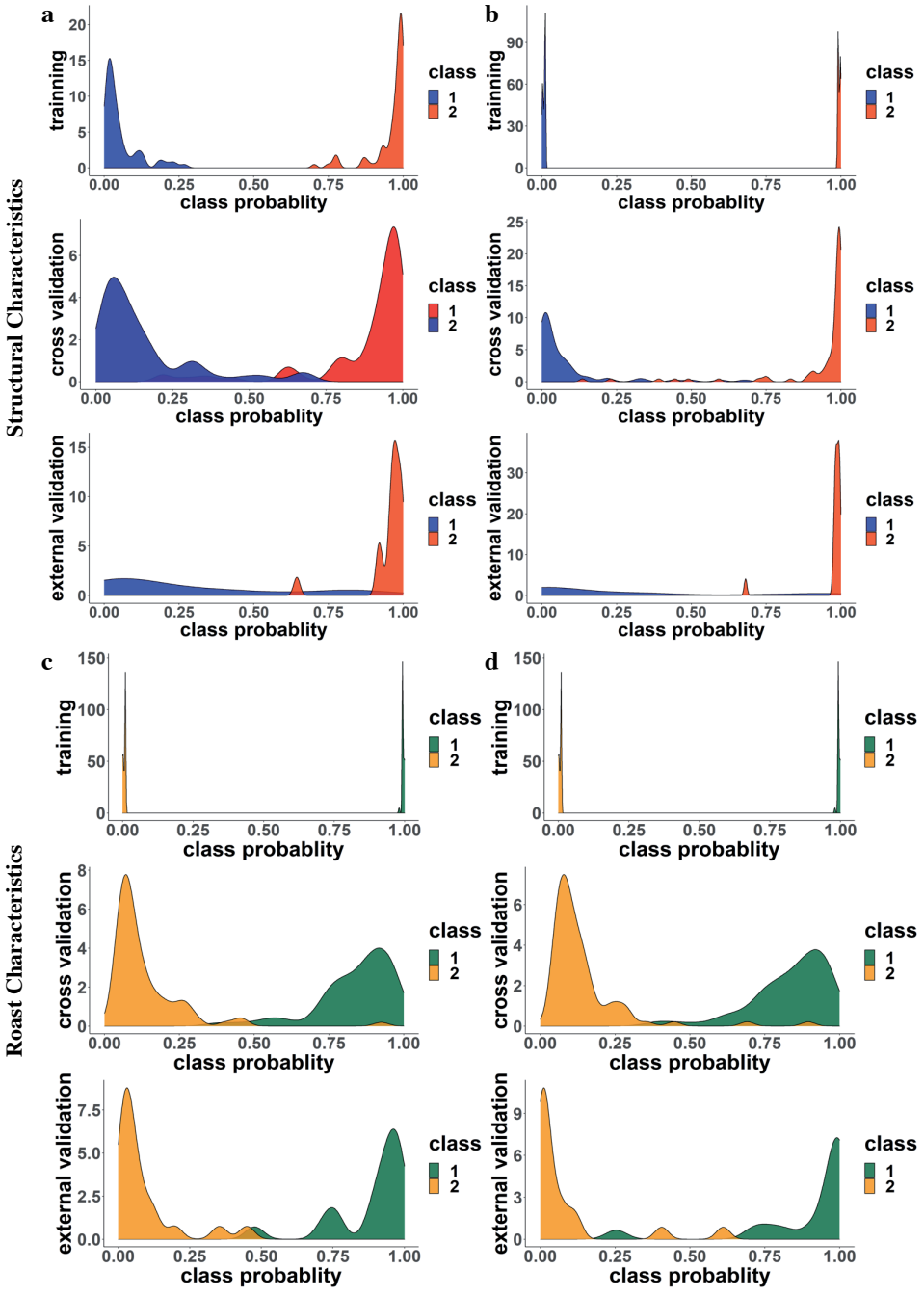


Figure 5.2. The kernel density estimation of probability distributions of different models, (a) FD-RF, (b) FD-RFSBF-SVM, (c) SNV-SVM, and (d) SNV-RFSBF-SVM. (FD) The first derivative; (RF) Random forest; (RFSBF) Random forest selection by filter; (SNV) Standard normal variate; (SVM) Support vector machine.

small peak in the wrong position. Despite this, SNV- RFSBF-SVM only used two parts of the full wavelength. These results provided further support for the hypothesis that features extracted could be used to simply models under the premise of ensuring the stability of model performances.

5.4. CONCLUSIONS

NIRS combined with various machine learning approaches was explored in this study to classify peanut varieties for precision processing based on the structure and roast characteristics of the resulting peanut butter. Cross validation, external validation, and KDE showed good performances of the models. To date, manufacturers in the peanut processing industry have solely used their inherent knowledge and experience to produce blends of peanuts to obtain peanut butter with different characteristics. The overall results of the study showed the feasibility of using NIRS to classify peanut varieties to product different peanut butters. This enables the possibility to sort or select peanut varieties based on the expected peanut butter qualities. Systematically scanning all peanuts could provide some objective data to predict the final product characteristics and thus reduce the waste of materials along the processing chain. Further work would be needed to increase the sample size and investigate the interactions of the processing conditions to provide guidance for adapted processing procedures to attain stable peanut butters.

REFERENCES

Ahmed, J., & Ramaswamy, H. S. (2006). Viscoelastic properties of sweet potato puree infant food. *Journal of Food Engineering, 74*(3), 376-382.

Alewijn, M., van der Voet, H., & van Ruth, S. (2016). Validation of multivariate classification methods using analytical fingerprints – concept and case study on organic feed for laying hens. *Journal of Food Composition and Analysis, 51*, 15-23. https://doi.org/10.1016/j.jfca.2016.06.003.

Baker, G., Cornell, J., Gorbet, D., O'keefe, S., Sims, C., & Talcott, S. (2003). Determination of pyrazine and flavor variations in peanut genotypes during roasting. *Journal of Food Science, 68*(1), 394-400. https://doi.org/10.1111/j.1365-2621.2003.tb14171.x.

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. https://doi.org/10.1016/j.foodres.2014.01.005.

Dhamsaniya, N. K., Patel, N. C., & Dabhi, M. N. (2012). Selection of groundnut variety for making a good quality peanut butter. *Journal of Food Science and Technology, 49*(1), 115-118. https://doi.org/10.1007/s13197-011-0361-9.

Faostat, F. (2020). Statistical databases. Retrieved from: http://www.fao.org/faostat/en/#data/QC Accessed.

Gong, A., Shi, A., Liu, H., Yu, H., Liu, L., Lin, W., & Wang, Q. (2018). Relationship of chemical properties of different peanut varieties to peanut butter storage stability. *Journal of Integrative Agriculture, 17*(5), 1003-1010. https://doi.org/10.1016/s2095-3119(18)61919-7.

Hourant, P., Baeten, V., Morales, M. T., Meurens, M., & Aparicio, R. (2000). Oil and fat classification by selected bands of near-infrared spectroscopy. *Applied Spectroscopy, 54*(8), 1168-1174. https://doi.org/10.1366/0003702001950733.

Kuhn, M. (2008). Building predictive models in R using the caret package. *J Stat Softw, 28*(5), 1-26.

Kuhn, M. (2012). Variable importance using the caret package. *Journal of Statistical Software, 6*.

Kuhn, M., & Johnson, K. (2013). An introduction to feature selection. In *Applied predictive modeling* (pp. 487-519): Springer.

Li, C., & Hou, L. (2018). Review on Volatile Flavor Components of Roasted Oilseeds and Their Products. *Grain & Oil Science and Technology, 1*(4), 151-156. https://doi.org/10.3724/sp.J.1447.Gost.2018.18052.

Liu, N., Parra, H. A., Pustjens, A., Hettinga, K., Mongondry, P., & van Ruth, S. M. (2018). Evaluation of portable near-infrared spectroscopy for organic milk authentication. *Talanta, 184*, 128-135. https://doi.org/10.1016/j.talanta.2018.02.097.

Liu, Y., Yu, Y., Liu, C., Regenstein, J. M., Liu, X., & Zhou, P. (2019). Rheological and mechanical behavior of milk protein composite gel for extrusion-based 3D food printing. *LWT - Food Science and Technology, 102*, 338-346. https://doi.org/10.1016/j.lwt.2018.12.053.

Mohd Rozalli, N. H., Chin, N. L., & Yusof, Y. A. (2015). Particle Size Distribution of Natural Peanut Butter and Its Dynamic Rheological Properties. *International Journal of Food Properties, 18*(9), 1888-1894. https://doi.org/10.1080/10942912.2014.971184.

Monforte, A. R., Martins, S., & Silva Ferreira, A. C. (2021). Discrimination of white wine ageing based on untarget peak picking approach with multi-class target coupled with machine learning algorithms. *Food Chemistry, 352*, 129288. https://doi.org/10.1016/j.foodchem.2021.129288.

Phan, Q., & Tomasino, E. (2021). Untargeted lipidomic approach in studying pinot noir wine lipids and predicting wine origin. *Food Chemistry, 355*, 129409. https://doi.org/10.1016/j.foodchem.2021.129409.

Resch, J. J., & Daubert, C. R. (2002). Rheological and Physicochemical Properties of Derivatized Whey Protein Concentrate Powders. *International Journal of Food Properties, 5*(2), 419-434. https://doi.org/10.1081/jfp-120005795.

Shakerardekani, A., Karim, R., Ghazali, H. M., & Chin, N. L. (2013). Textural, rheological and sensory properties and oxidative stability of nut spreads–a review. *International Journal of Molecular Sciences, 14*(2), 4223-4241. https://doi.org/10.3390/ijms14024223.

Song, W., Wang, H., Maguire, P., & Nibouche, O. (2018). Collaborative representation based classifier with partial least squares regression for the classification of spectral data. *Chemometrics and Intelligent Laboratory Systems, 182*, 79-86. https://doi.org/10.1016/j.chemolab.2018.08.011.

Sun, A., & Gunasekaran, S. (2009). Yield Stress in Foods: Measurements and Applications. *International Journal of Food Properties, 12*(1), 70-101. https://doi.org/10.1080/10942910802308502.

Sundaram, J., Kandala, C. V., & Butts, C. L. (2009). Application of near infrared spectroscopy to peanut grading and quality analysis: overview. *Sensing and Instrumentation for Food Quality and Safety, 3*(3), 156-164. https://doi.org/10.1007/s11694-009-9081-5.

Wang, L., Wang, Q., Liu, H., Liu, L., & Du, Y. (2013). Determining the contents of protein and amino acids in peanuts using near-infrared reflectance spectroscopy. *Journal of Food Science and Technology, 93*(1), 118-124. https://doi.org/10.1002/jsfa.5738.

Wang, Q. (2016). *Peanuts: Processing technology and product development*. Cambridge: Academic Press.

Wang, Q. (2018). *Peanut processing characteristics and quality evaluation*. New York: Springer.

Wang, Q., Liu, H., Shi, A., Hu, H., Liu, L., Wang, L., & Yu, H. (2017). Review on the processing characteristics of cereals and oilseeds and their processing suitability evaluation technology. *Journal of Integrative Agriculture, 16*(12), 2886-2897. https://doi.org/10.1016/s2095-3119(17)61799-4.

Windham, W., Kandala, C., Sundaram, J., & Nuti, R. (2010). Determination of peanut pod maturity by near-infrared reflectance spectroscopy. *Transactions of the ASABE, 53*(2), 491-495.

Yan, J., Stuijvenberg, L., & 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). https://doi.org/10.1002/ejlt.201900031.

Yu, H., Liu, H., Erasmus, S. W., Zhao, S., Wang, Q., & van Ruth, S. M. (2020). Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics. *Industrial Crops and Products, 158*. https://doi.org/10.1016/j.indcrop.2020.112956.

Yu, H., Liu, H., Erasmus, S. W., Zhao, S., Wang, Q., & van Ruth, S. M. (2021). An explorative study on the relationships between the quality traits of peanut varieties and their peanut butters. *LWT - Food Science and Technology, 151.* https://doi.org/10.1016/j.lwt.2021.112068.

Yu, H., Liu, H., Wang, N., Yang, Y., Shi, A., Liu, L., Hu, H., Mzimbiri, R. I., & Wang, Q. (2016). Rapid and visual measurement of fat content in peanuts by using the hyperspectral imaging technique with chemometrics. *Analytical Methods, 8*(41), 7482-7492. https://doi.org/10.1039/c6ay02029a.

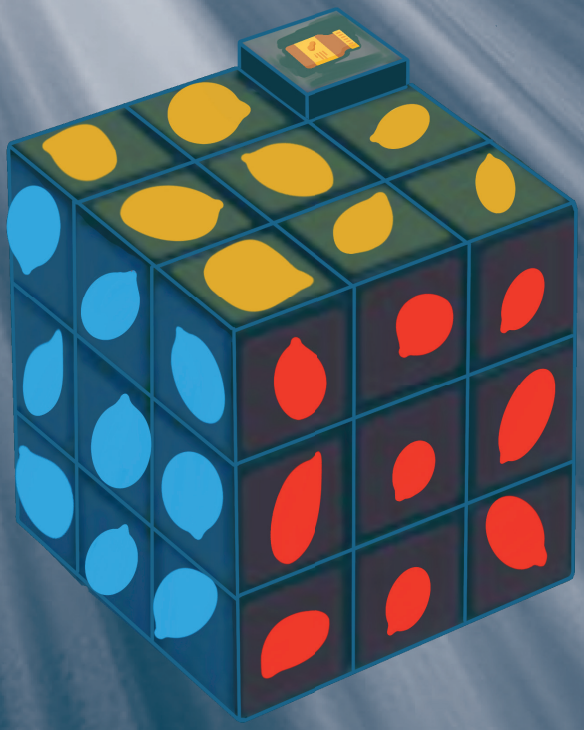
Yu, H., Liu, H., Wang, Q., & van Ruth, S. (2020). Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. *LWT - Food Science and Technology*, 109398. https://doi.org/10.1016/j.lwt.2020.109398.

Zhao, S. M., Yu, H. W., Gao, G. Y., Chen, N., Wang, B. Y., Wang, Q., & Liu, H. Z. (2021). Rapid Determination of Protein Components and Their Subunits in Peanut Based on Near Infrared Technology. *Guang Pu Xue Yu Guang Pu Fen Xi/Spectroscopy and Spectral Analysis, 41*(3), 912-917. https://doi.org/10.3964/j.issn.1000-0593(2021)03-0912-06.

SUPPLEMENTARY MATERIALS

Table S5.1 The detailed information of peanut varieties.

Number	Sample name	High oleic acid peanut
1	Zhongkaihua1	No
2	Zhanyou75	No
3	Yueyou7	No
4	Yuhua9326	No
5	Yuhua76	Yes
6	Yuhua65	Yes
7	Yuhua37	Yes
8	Weihua8	No
9	Weihua25	Yes
10	Weihua16	No
11	Tianfu18	No
12	Shanhua9	No
13	Shanhua8	No
14	Shanhua7	No
15	Shanhua11	No
16	Puhua28	No
17	Luhua11	No
18	Jinhua9	No
19	Jinhua8	No
20	Jihuatian1	No
21	Jihua8	No
22	Jihua5	No
23	Jihua19	Yes
24	Jihua18	Yes
25	Jihua16	Yes
26	Jihua10	No
27	Huayu963	Yes
28	Huayu962	Yes
29	Huayu917	Yes
30	Huyua661	Yes
31	Huayu33	No
32	Huayu25	No
33	Huayu22	No
34	Huayu20	No
35	Hanghua2	No
36	Fuhua12	No
37	Fenghua1	No
38	Baisha1016	No
39	Baisha	No
40	DF06	Yes



CHAPTER 6

General discussion

6.1. OVERVIEW

The quality traits of peanuts are the foundation of the entire peanut industry with the globalisation of the peanut industrial chain. Peanut quality traits attract more and more attention which makes that the demands for rapid quality traits evaluation of peanuts and their products are increasing. Therefore, this thesis is to elucidate and comprehend distinct analytical signatures and relationships of various types of peanuts and derived peanut butters for quick evaluation and identification of peanuts to improve the values of peanuts and final products in the whole production chain. To achieve this aim, **Chapters 2** and **3** examined the different analytical signatures of batch samples and single kernel peanuts, and rapid detection methods based on portable near-infrared spectroscopy (NIRS) were developed and evaluated (Yu, Liu, Erasmus, et al., 2020; Yu, Liu, Wang, et al., 2020). Next to that, the relationship between peanuts and peanut butters was investigated in **Chapter 4** (Yu et al., 2021). In **Chapter 5**, the rapid methods of peanut classification for peanut butter manufacture were developed and evaluated. In the current chapter, the findings from **Chapters 2-5** are presented in **Table 6.1**, along with the discussion and interpretations of the main findings. Following this, different analysis methods are compared and the implications for the industry are evaluated. Finally, the limitations and recommendations for future research are provided.

6.2. THE ANALYTICAL SIGNATURES OF DIFFERENT PEANUT VARIETIES AND THE ESTABLISHMENT OF THE SCREENING METHODS

6.2.1. The analytical signatures of different peanut varieties

In this thesis, the major components' profiles of different peanut varieties were studied in **Chapters 2-4**. Specifically, the fatty acids (FAs) profiles of high oleic acid peanuts (HOP) and regular peanuts (RP) were investigated in **Chapter 2**. The content of oleic acid was significantly higher ($P < 0.001$) in HOP, while the contents of linoleic acid and palmitic acid were significantly lower ($P < 0.001$) than in RP. In **Chapter 3**, the major analytical signatures (e.g. fat, protein, sucrose, and amino acids) of different peanut varieties excluding FAs were evaluated. The principal component analysis (PCA) biplot (**Figure 3.2**) showed that fat and sucrose are the most important components that distinguish peanut varieties. When adding FAs, however, FAs appear to be even more important (**Figure 4.2**) to distinguish peanut varieties. That is because the FAs composition differs to a larger extent across varieties than the other characteristics. The main three reasons for the differences and similarities in the analytical signatures between varieties were gene factors, growing environment, and agricultural practices. For instance, *ahFAD2* gene families are the feature genes of HOP (Nawade et al., 2018). The growing environment (such as soil and

Table 6.1. The summary of the main findings described in this thesis.

Chapter	Aim	Main findings
2	To investigate the differences of FAs between HOP and RP in order to classify and quantify its major FAs by portable NIRS, and compare results with benchtop NIRS.	• The oleic acid, linoleic acid, and palmitic acid were highly significant ($P < 0.001$) between HOP and RP and the interrelationships between them were determined by PCA. • HOP and RP were completely distinguished based on PLS-DA by portable and benchtop NIRS. • Portable NIRS had the same quantitative performance for major FAs prediction as benchtop NIRS in terms of PLS method.
3	To explore the differences and similarities of fat, sucrose, protein, and amino acids between different peanut varieties in order to establish quantitative prediction models of the above compounds at the single peanut level by the portable NIRS combined with the single peanut detection accessory.	• The compositional patterns of different samples were clustered into four categories explored by PCA combined with K-means method. • A detection accessory of single peanuts was designed to collect spectral data. • The spectral data was explained by the structure of molecular groups of the peanuts. • The quantitative analysis of single peanuts based on PLS was implemented by the portable NIRS.
4	To explore the variances of the quality traits of peanut butters from different peanut varieties and elucidate their multivariate relationships.	• Broad variations in the quality traits of peanuts and peanut butters were found. • The quality traits patterns of different peanut butters could be separated into three distinct groups. • Group 1 had the highest values of pyrazines, texture, and rheological properties. • Peanuts with higher sucrose content and lower fat content had great contributions to the texture, rheology, and pyrazines of peanut butters.
5	To conduct cluster analyses of peanut varieties based on the quality traits analysis results of the corresponding peanut butters in order to rapidly identify the characteristics of peanut butters through the acquired NIRS data of peanuts.	• Significant differences ($P < 0.05$) were found in the structure and roast traits of peanut butters between different groups. • The first derivative with random forest was the best method for structure traits identification model. • Standard normalisation variable with support vector machine was the best method for roast traits identification model. • Random forest selection by the filter can effectively select the feature variables to simplify the models.

FAs, Fatty acids; HOP, High oleic acid peanuts; NIRS, Near-infrared spectroscopy; PCA, Principal component analysis; PLS, Partial least squares; PLS-DA, Partial least squares discriminant analysis; RP, Regular peanuts.

climatic factors) and agricultural practices (such as mix cropping and weed management) as the main externalities have great influences on the quality traits of peanuts (Chamberlin, 2019; Phan-Thien et al., 2010; Zuo & Zhang, 2008). In terms of physical analytical signatures of peanuts, such as kernel length, kernel width, and kernel weight, there is no clear relationship between the physical characteristics and chemical characteristics. Therefore, according to the above results, peanut variety is key, especially when it comes to the high oleic acid content. Following this, fat and sucrose are used to characterise peanut varieties for further utilisation. Basically, peanut varieties with high fat content always have low

sucrose content (Bishi et al., 2013; Yu, Liu, Erasmus, et al., 2020). Meanwhile, we should consider protein and amino acids contents between different peanut varieties as well, especially arginine which is the most abundant amino acid compared with other cereals and crops (Wang, 2016).

6.2.2. The establishment of the screening methods

Portable NIRS and benchtop Fourier transform NIRS were utilised to explore the characteristic molecular vibrations of different peanut varieties in **Chapters 2** and **3**. Even in a similar wavelength range, the spectral data collected by the benchtop NIRS have lower and concentrated reflectance ranges (e.g. 0.26 - 0.33 at the starting position) compared with portable NIRS (**Figure 2.3**). The distinctions may be explained by the differences of the light path inside the instruments and the sample cup size. The light path literally makes the reflection value higher or lower (Liu et al., 2018; Xu et al., 2018), and the concentrated extent of the reflection spectra depends on the diffuse reflection degree (Rady et al., 2019). The shape of peanut kernels is mainly oval, leading to diffuse reflection of light. The larger size sample cup used with the benchtop NIRS avoids too much space between peanut kernels which reduces the diffusion effect (Williams, 2021). Meanwhile, the spectral characteristics of batch and single kernel measurements by portable NIRS are fairly similar (**Figure 2.3** and **Figure 3.5)**. The observed differences are that the absorbance values of the spectral data of single kernels are lower than for batch measurements. The absorbance value equals the -log transformation of the reflectance value (Gardner, 2018). Compared with single kernels scanning which has the largest scanning area, batch scanning probably has a partial scanning area because of the gap between peanuts caused by the samples' loading. The distinctions in the scanning area result in higher reflection values for single peanut kernels and then lower absorbance values.

Although there are differences between portable analyses (batch detection and single kernel detection) and benchtop analyses (batch detection), the spectral data of the three types show similar trends because they are both based on the same molecular vibrations. For the raw spectral data, the peaks and valleys are 1106 nm, 1205 nm, 1298 nm, and 1460 nm, which are features for peanuts compared with other crops such as soybean (Kusumaningrum et al., 2018) and wheat (Shi & Yu, 2017). After pre-treatments, such as the second derivative, more peaks appeared involving 927 nm, 952 nm, 1007 nm, 1038 nm, 1125 nm, 1162 nm, 1181 nm, 1212 nm, and 1243 nm. The molecular vibrations are related to the bonds of C-C, C=C, N-H, O-H, etc., which indicate the concentrations of some compounds, including fat, FAs, protein, amino acids, etc. Theoretically, the band locations linked with fat and FAs are identified at 930-960 nm as the third vibration and the fourth stretching overtone of C-H group (Workman & Weyer, 2012; Yu et al., 2016). N-H stretch second and first overtone principally originate from protein and amino acids,

which contribute to 973-1020 nm and 1420-1530 nm of the spectral structure, respectively (Workman et al., 2012). In addition, O-H and C-H stretch mainly stemmed from sucrose lead to 1400-1450 nm of the spectral structure (da Costa Filho, 2009; Workman et al., 2012). Hence, the features of NIRS in certain wavelength ranges could be treated as the derived characteristics of the different compounds. As a result, the classification models have high performances to distinguish HOP from RP based on the spectral data linked with different chemometrics approaches by the portable NIRS device. Furthermore, fat content, individual FAs, sucrose content, protein content, and amino acids could be reliably and quantitatively predicted from the spectral data.

6.3. THE RELATIONSHIPS BETWEEN PEANUTS AND DERIVED PEANUT BUTTERS

In order to elucidate the relationship between peanuts and peanut butters, the characteristics of different peanut varieties and derived peanut butters were analysed in **Chapters 4-5**. Moreover, the similarities and distinctions in quality traits between the different groups were assessed. Raw materials were grouped into three clusters based on the physical and chemical qualities of peanuts, especially the contributions of fatty acids (**Figure 4.2)**. Similarly, the derived peanut butters were also grouped into three clusters based on the structure characteristics (texture and rheology) and roast characteristics (colour and volatile compounds) together, where the groups had significant differences (**Figure 4.4**). Furthermore, there is no significant correlation between the structure characteristics and roast characteristics (**Figures 4.3b** and **4.5b)**. Hence, the quality traits of peanut butters could subsequently be divided into two categories: the texture and rheological characteristics reflecting the internal structure of peanut butters, and colour and volatile compounds reflecting the roast characteristics. The new cluster results were presented in **Figure 5.1 a** and **b**.

The separate analyses combined with the overall analysis provided a deeper insight that the components of raw material profoundly influenced the quality traits of peanut butter. For instance, the structure characteristics of peanut butter are determined through interactions between major components, such as fat, protein, and carbohydrates. Peanut butter is oil-based colloidal suspension, where solid flecks are diffused. Fat, accounting for above 50% of total mass, is the lubricant for the whole system to mitigate the interactions, while protein and carbohydrates act like thickeners to boost the internal structure of peanut butter. It has been underlined that fat content had a significant ($P < 0.05$) negative relationship with the firmness of peanut butters in this thesis. Meanwhile, the compositions of fat, especially FAs profiles, are not negligible factors for the structure characteristics because the ratio of unsaturated to saturated FAs was significantly correlated with struc-

ture characteristics (Staniewski et al., 2021). Although this thesis showed that oleic acid had a positive relationship with rheological characteristics while linoleic acid and palmitic acid had negative relationships with rheological characteristics, they were not significant. The latent cause was that oleic acid structures, especially double bonds, were destroyed because peanut butter preparation had gone through high temperature (the roast processing) and high-speed cutting (the grinder processing). Nonetheless, peanut butters manufactured with high HOP have a longer shelf life (Gong et al., 2018) and nutritional benefits (Zhao et al., 2019). The manufacturers could select HOP varieties from group 1 (e.g. JH18, YH37, and HY963) recommended in this thesis to increase the competitiveness of peanut butters through texture and flavour. In contrast, protein and sucrose have significant ($P < 0.05$) positive relationships with structure characteristics of peanut butters. Specifically, the average diameter of peanut protein particles are 2-10 μm, which gives peanut butter a soft and thick texture (Damodaran et al., 2007). Therefore, an increase in protein content literally boosts the internal structure of peanut butter. Meanwhile, sucrose is one of the key components which influenced the texture and rheological characteristics. It is known that sucrose could lead to an increase in the macromolecular entanglement (Mathlouthi & Reiser, 2012). As a result, there are significant differences ($P < 0.05$) in the structure characteristics between two groups of peanut varieties in **Chapter 5**.

In terms of the roast characteristics, non-enzymatic browning, especially the Maillard reaction, results in the colour and volatile compounds of peanut butters. In this thesis, the roast condition was consistent with a temperature of 160°C and a time of 30 min. Peanut varieties were the single variable for the Millard reaction. On one hand, different compositions of peanut varieties influence the Maillard reaction rate. For instance, in addition to pentose, the effect of hexose (glucose and fructose) on the rate is greater than that of other monosaccharides. Although sucrose as non-reducing sugar cannot be directly involved in the Maillard reaction, sucrose can hydrolyse into glucose and fructose during the heating process (Plazl et al., 1995). The Maillard resulting volatile products such as pyrazines formed in the final phase were shown to be affected by decreasing sucrose content (Garvey et al., 2021). Similarly, glycine is the most active amino acid for the Maillard reaction (Wang & Qi, 2015). Therefore, sucrose and glycine showed a significant ($P < 0.05$) positive correlation with volatile compounds. On the other hand, the compositional contents directly affect the formation of volatile compounds, such as serine, histidine, and leucine for pyrazines, and monosaccharides for pyrrole. This is because amino acids (nitrogen supply) and saccharides (carbon sources) are two key precursor substances (Hui et al., 2010).

In **Chapter 5**, we demonstrated the hypothesis that the spectral data of raw materials could predict the quality traits of peanut butters. Theoretically, the spectral data of peanuts could reflect the similarities and differences in the chemical compositions of peanut

varieties shown in **Chapters 2-3**. The relationship between the chemical compositions of peanuts and the quality traits of peanut butter has been analysed **in Chapter 4**. The underlying logic is obvious that the chemical composition of peanut varieties bridges the spectral data of peanuts and the quality traits of peanut butters. Similar results have been found in previous studies, e.g. in regard to the spectral data of apples for predicting the qualities of purees (Lan et al., 2020), and the spectral data of wheat varieties for predicting the qualities of chapatti (Kundu et al., 2017). According to the results of **Figure 5.1 c-g**, the spectral data of raw materials showed the differences between different peanut butter groups, especially the wavenumber range from 8765.62 cm^{-1} to 3996.02 cm^{-1}. These wavenumber ranges are derived from the key components that affect the quality traits of peanut butters.

Compared with partial least squares discrimination analysis, machine learning, such as support vector machine and random forest, have a greater capability to deal with the nonlinearity and collinearity of the multivariable to improve the performances of models established by the full wavelength range and feature wavelengths in this thesis. Machine learning advantages in handling complex situations were also found in a previous study (Deng et al., 2020). This thesis offered reliable methods to evaluate the suitability of peanut manufacturing peanut butter. What needs to be improved is the specificity and sensitivity of the models due to the limitations of the number of samples. In practice, the models will be more refined as more samples are added and new methods are introduced such as gradient boosting and neural networks.

6.4. EVALUATION OF THE LAB-BASED CONFIRMATORY AND SCREENING METHODS

In the current thesis, the quality traits of peanuts and peanut butters were analysed by the lab-based confirmatory methods (e.g. gas chromatography for FAs, amino acid automatic analyser, high performance anion exchange chromatography for sucrose, gas chromatography-mass spectrometry (GC-MS) for the volatile compounds of peanut butters, etc.) and screening methods (portable NIRS and benchtop NIRS). To compare these analytical methods, several aspects should be considered, which include sample preparation, operational difficulty, analysis time, the price of the instruments, operational environment, and precision performance (Müller-Maatsch, Bertani, et al., 2021).

Sample preparation. The sample pre-treatments for the lab-based confirmatory test and screening test vary significantly. According to the standard methods for the compositional analysis in peanuts and the chemical methods for the quality analysis in peanut butters, chemical reagents must be applied to perform arduous sample pre-treatments.

Some of the pre-treatments especially are demanding in terms of high temperature and long time. For example, peanuts need to be hydrolysed at 70°C and methylated at 100°C before FAs analysis (Anderson et al., 2019). Similarly, amino acids analysis requires samples to be hydrolysed at 110°C for 22 h. Therefore, the sample preparation of lab-based confirmatory methods is complicated, time-consuming, and unsustainable. In contrast, the protocol of the screening methods such as portable NIRS and benchtop NIRS is simple and green even though the initial phase of model establishment requires the lab-based confirmatory test as a reference, which means that the quality traits of peanuts can be detected directly without any sample preparation (Fernandez-Novales et al., 2019).

Operational difficulty. In terms of the lab-based confirmatory analyses, both sample pre-treatments and detections have strict requirements on the manual operation and the parameters settings (Yan et al., 2019). In particular, the operation of a GC-MS system needs well-trained technicians. Meanwhile, professional training is required to adopt the advanced software from data collection to analysis. Overall, the operational process requires excessive time and professional personnel. However, regarding portable NIRS and benchtop NIRS, the method establishment and application can be separated. Hence, although the method establishment needs to use various data, chemometrics and associative knowledge, and expertise level, it is relatively straightforward to apply. The whole protocol specifies the basic operation to reduce the noise including spectral device preheating and calibration, and sample loading and measurement. Accordingly, untrained personnel could be tutored quickly to master the protocol because the operation is easy and the software has the friendly graphical user interface, while professional technicians are only required to advance the models.

Sample analysis time. According to the sample preparation and operation, the lab-based confirmatory analyses always take considerable time. Specifically, it takes more than 10 h per sample to separate and analyse by GC-MS for the FAs analysis (Yu, Liu, Wang, et al., 2020). Similarly, the protocols of the major components and amino acids analysis expect a significant amount of time to obtain the corresponding results (Yu, Liu, Erasmus, et al., 2020). As regards the quality traits of peanut butters, a 4-hour analytical process was adopted to identify volatile compounds (Yu et al., 2021). Even if most equipment includes automated systems, it is inevitably to occupy a long time to fulfil the whole analyses. In the case of spectral analysis, it naturally needs to spend enough time in creating substantial databases to establish models. But once the methods are established, the regular analysis time per sample is about 1 min including sample preparation, experimental operation, and sample analysis. More importantly, if the instrument involves abundant quality traits models, such as fat, protein, and amino acids, it can detect the quality traits of peanuts simultaneously. Therefore, the efficiency of spectral analysis is further improved and it would be optimal when considering sample preparation and analysis time.

Laboratory or on-site test. The lab-based confirmatory analyses, such as using GS-MS, can only be performed in a laboratory since they usually require multi-step, high temperature, special apparatus, and rigorous conditions. The application scenarios of benchtop NIRS are also limited to the laboratory due to its sophisticated optical components. Miniaturised spectroscopic devices have been developed to replace the large and immobile analytical instruments and can be applied in situ (Müller-Maatsch & van Ruth, 2021). In this thesis, the commercial miniaturised spectroscopic device is remodelled for peanut detection. The device is equipped with an aviation material cabinet of 420 mm×330 mm×166 mm size, which is easy to carry and transport anywhere (Müller-Maatsch, Bertani, et al., 2021). The equipment has two detection accessories for batch measurements and single kernels detections. At the same time, it has an inline battery for continuous power supply for 10 h to guarantee use for one day. In summary, portable NIRS has more advantages than GC-MS and benchtop NIRS regarding in situ applications.

Price of the instrument. Pricing is a critical consideration for enterprises and relevant government departments, when it comes to the selection of analytical equipment. The small companies commonly cannot afford the cost of the lab-based confirmatory analysis instruments which are always purchased by large enterprises, research institutions, and commercial laboratories because they are worth more than €50,000. Meanwhile, these instruments need irregular maintenance and replacement of components after use, which increase the total cost. Both portable NIRS and benchtop NIRS are huge competitive at this point because their prices are obviously lower than the lab-based confirmatory test instruments. Especially, the price of portable NIRS is less than €10,000. In terms of maintenance charges, the main cost is to change the light sources but the light sources generally have a long service life. With the development of the miniaturised spectral instruments, the market pricing of instruments preparation is gradually decreased. Therefore, portable NIRS can serve more clients without huge financial stress.

Precision performance. Although the screening methods have many merits, they cannot completely replace the lab-based confirmatory methods because the performances of the screening methods are affected by several factors. As indicated previously, the establishment of the screening methods hinges on the lab-based confirmatory methods as references. Taking NIRS methods as examples, these factors include the number and variance of representative samples, the accuracy of the reference measurements, and the stability and operational environment of the device. Hence, once deviations from expected are found after the screening, all the abnormal samples should be transferred to the laboratory for further confirmation by the lab-based confirmatory methods. In addition, these abnormal samples could be added to the original sample set to update the models.

6.5. IMPLICATIONS FOR PEANUT INDUSTRY

To improve the values of products and economic benefits in the whole peanut chain, the screening methods (NIRS) could be used to determine key quality traits of peanuts at three key points in the chain: breeding, trade, and processing (**Figure 6.1**). Formerly, the traditional approaches in each point were the chemical methods which are destructive and non-efficient. With the whole development of the peanut industry, from breeders to farmers, from processors to regulators, all stakeholders pay more attention to the quality traits of peanuts and their products. Currently, with the screening methods proposed, the whole peanut chain will benefit because the quality traits of peanuts can be quickly and accurately assessed. The three points are detailed below.

Single peanut kernels assessment. Breeding is the starting point of the whole peanut industry. Plant breeding experts are always seeking approaches to enhance the quality traits of current varieties since a good peanut variety can bring huge economic benefits to farmers and improve the values of products for consumers. During the breeding process, only a few kernels have genetic mutations to attain the desired traits because the heritability of the targeted trait might be below. As a result, each generation of seed resources is limited and valuable. Therefore, it is difficult to analyse the quality traits of seed resources by the conventional benchtop NIRS which generally needs about 250 g samples and the chemical methods which are destructive (Agelet & Hurburgh, 2014). A single peanut kernel NIRS analyser can offer non-destructive, rapid, and high throughput detection methods, which boost the breeding process and protect rare peanut resources. The corresponding quantitative approaches of main components such as fat, protein, sucrose, and amino acids have wide applications.

On-site screening methods for trade. After the harvest of mature peanuts, peanuts will be traded to domestic and international companies for further manufacture. The peanut grade as the focus of the trade, which is mainly determined by the compounds such as fat, FAs, and protein, is critical to ensure a healthy trading market. Meanwhile, the transactions could take place in various sites from the fields to the customers. The usual practice is to send samples to a laboratory for testing. The long waiting for results would probably miss the perfect trading opportunity. On the contrary, portable NIRS can offer better detection with cost-efficient and fast results. Hence, companies can quickly assess peanut authenticity (e.g. HOP) and quality traits to expedite transactions with a guarantee. Similarly, law enforcement agencies can take advantage of portable NIRS to supervise the entire supply chain to sustain market order. If the analysis findings cause suspicions, samples should be sent to the laboratory for further investigation by confirmatory analysis.

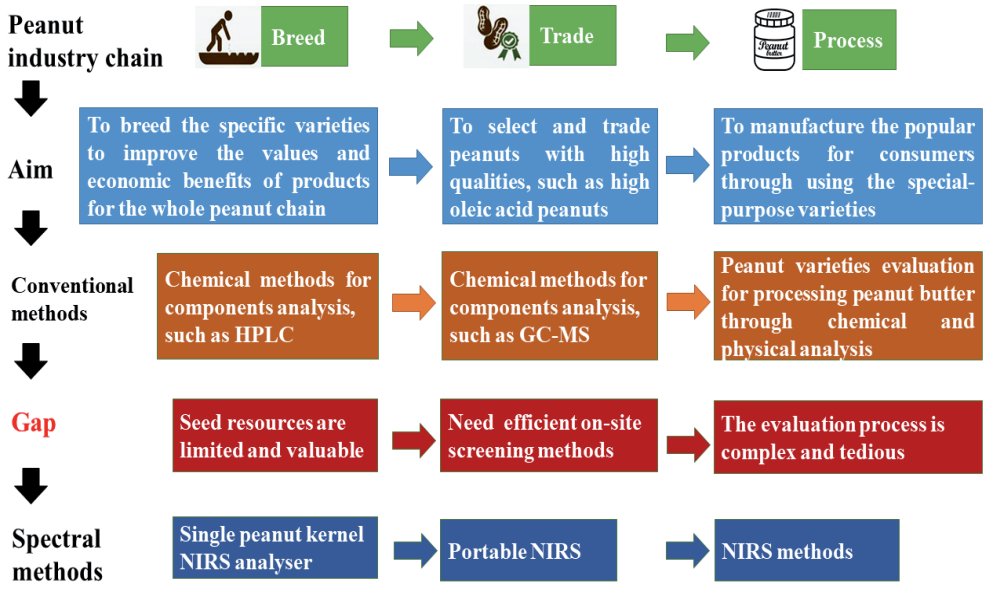


Figure 6.1. The application of the advanced detection techniques in the peanut industry chain. (HPLC) High performance liquid chromatography; (GC-MS) Gas chromatography–mass spectrometry; (NIRS) Near-infrared spectroscopy.

Selection of peanut varieties for manufacture with purpose. The quality traits of final peanut products are basically influenced by two major important factors: raw materials and processing conditions. Just as different flour types require distinct wheat varieties (e.g. protein quantity and quality) (Ortolan & Steel, 2017), peanut varieties, as the first workshop for processing, are also evaluated from the products' perspective. A previous study was conducted to classify peanut varieties for processing peanut protein with high gelation or high solubility (Wang et al., 2017). In this thesis, we classified peanut varieties for manufacturing different types of peanut butter. Meanwhile, the methods of the rapid assessment of peanut varieties for processing peanut butter were proposed. The companies can quickly choose peanut varieties for the different demands of consumers. In other words, they no longer need to manufacture peanut butter and measure the properties one by one to judge the suitability of peanut varieties as before. The proposed approaches are intended to enhance peanut foods' quality control, save expenses, and enable faster decision-making.

6.6. GENERAL CONCLUSIONS

The aim of this PhD thesis was to elucidate and comprehend distinct analytical signatures and relationships of various types of peanuts and derived peanut butters for quick

evaluation and identification of peanuts. According to the results described in the previous chapters, the following conclusions can be drawn.

It was found that different peanut varieties generally have various physical and chemical qualities (**Chapters 2** and **3**). Specifically, the FAs profiles differed significantly between the HOP varieties versus the RP varieties. Furthermore, when comparing the instrumental performance of portable NIRS and benchtop NIRS, the same performance was obtained for identifying peanut variety and for the detection of their associated FAs profiles. On the other hand, the fat, protein, sucrose, and amino acid composition of single kernel peanuts can be quickly detected by portable NIRS equipped with the single detection accessory and exploiting the differences in spectral data.

The impact of raw materials on the quality traits of peanut butters is considerable. A new classification method has been proposed from the perspective of peanut processing and utilisation. Peanut varieties are also classified into the same category after manufacturing into peanut butters. The underlying reason is that the related chemical compositions (e.g. fat, sucrose, and amino acids composition) of peanut varieties in the same group are comparable. Meanwhile, one of the major findings to emerge from the thesis is that the spectral data of the raw materials can predict the quality traits of peanut butters, and reversely, the processed raw materials can be inferred based on the quality traits of peanut butters. Therefore, NIRS can rapidly distinguish peanut varieties to predict the properties of the expected peanut butters, and potentially peanut butters could be checked for their raw materials after processing.

6.7. RESEARCH LIMITATIONS AND RECOMMENDATIONS

The reliable theoretical basis and rapid detection methods in this thesis will facilitate to monitor, adjust, and guarantee the quality and identity of peanuts in the whole production chain. Meanwhile, peanuts and their product processing are globalised and complex, leading to many quality and safety issues (Wang, 2018). Some issues happening in the peanut industry could be accidental or intentional. Nevertheless, there is always a need for rapid quality and safety analysis, especially spectroscopy techniques. However, there are still some limitations to this thesis. In order to improve the performances of the models and have a deeper insight into the whole peanut chain, some suggestions are proposed below.

Sample size and models updated. The establishments of NIRS detection methods rely on the reference samples and reference values used. Therefore, the quantity and representation of samples are indispensable to ensure the robustness of the models. On one

hand, new peanut varieties will be cultivated every year. These varieties are unknown to the current models. On the other hand, the prevailing varieties should consider the variances in the environmental factors (e.g. year, soil, and climate) and farm system (e.g. planting pattern and management). Meanwhile, the peanut samples from other main peanut planting countries such as India and the USA should be collected to advance models for international trade. With the number of the above samples added and leveraged, our methods could have more adaptability to be operated worldwide.

Processing conditions of peanut butter and the qualities prediction. Processing conditions are the other determinant factor in addition to the raw materials. The processing conditions remained unchanged in order to control variables in this thesis. However, for the same variety, the quality traits of peanut butter would be diverse under different processing conditions. For example, the roast temperature and time have profound influences on the types and contents of volatile compounds. Processing good peanut butters not only require some typical peanut varieties but also utilise the optimum production process conditions. Therefore, in the future, the relationship between different process conditions and the characteristics of peanut butter could be explored to further improve the product qualities. Meanwhile, the characteristics of peanut butter should be measured directly by NIRS through chemometrics to link the spectral data of peanut butter. The potential application is that it could help to assess the qualities of peanuts butter and discriminate peanut butter adulteration (e.g. corn flour, pear seed, and cassava flour).

The instruments for the whole peanut chain. Currently, portable NIRS and benchtop NIRS are the main instruments used on-site and in the laboratory. From the perspective of the whole chain, there are still several demands for other types of rapid instruments. Firstly, the drone or ground-based lidar sensor could be used to monitor the growing conditions of peanut planting. For instance, the chlorophyll contents of leaves and the soil situation could be quickly obtained to pledge the sustainable development of the local planting (Qi et al., 2021; Yuan et al., 2019). Secondly, online detection devices such as hyperspectral imaging devices or optical fibre sensors should be considered to ensure process quality control. Manufacturers could control the vital links of production to guarantee the quality and safety of the peanut products by surface, subsurface, and internal evaluation (Garrido-Varo et al., 2018; Yu et al., 2016). Finally, the phone-based device should be designed for consumers (Hussain & Bowden, 2021). Smart phones are becoming more and more popular around the world. They are not solely communication instruments but have supercomputing power chips, large storage capacity, and long battery life. That means that smart phones could develop forwards a consumer-oriented device, and consumers could have a deep insight into the qualities of peanuts and their products before final purchase. Therefore, advances in high-performance hardware will drive the future development of the peanut industry.

Advanced chemometrics and Internet of Things. The use of chemometrics makes spectroscopy techniques more convenient and effective for the analysis of quality traits in peanuts and their products. However, the robustness of some models needs to be improved by new pre-treatments and algorithms. For instance, artificial neural networks and reinforcement learning can improve the performances of qualitative and quantitative models. Agricultural Internet of Things can integrate sensing, spectral, computing, and implementing devices to support need-specific operations (Tzounis et al., 2017), so that peanut producers, processers, and other stakeholders can properly and promptly respond to the situation of improving the yield and the quality traits of peanut and its products.

The quality traits of peanuts are the cornerstone of the entire industry. Although there are still some restrictions that limit these techniques for peanut industrial utilisation, it is anticipated that the increasing requirements for the quality and safety analyses in peanut and its products, as well as ongoing work on the development of spectroscopy methods, will make these techniques more effective for peanut industrial applications. This thesis tackled the barriers regarding the analytical signatures in batch and single kernel measurements as well as the relationship between peanuts and derived peanut butters. Meanwhile, the establishment of novel spectroscopic approaches makes the evaluation of peanuts more efficient. These improvements could offer a fresh perspective and spark additional research in the future.

REFERENCES

Agelet, L. E., & Hurburgh, C. R., Jr. (2014). Limitations and current applications of Near Infrared Spectroscopy for single seed analysis. *Talanta, 121*, 288-299. https://doi.org/10.1016/j.talanta.2013.12.038.

Anderson, J. V., Wittenberg, A., Li, H., & Berti, M. T. (2019). High throughput phenotyping of Camelina sativa seeds for crude protein, total oil, and fatty acids profile by near infrared spectroscopy. *Industrial Crops and Products, 137*, 501-507. https://doi.org/10.1016/j.indcrop.2019.04.075.

Bishi, S. K., Kumar, L., Dagla, M. C., Mahatma, M. K., Rathnakumar, A. L., Lalwani, H. B., & Misra, J. B. (2013). Characterization of Spanish peanut germplasm (Arachis hypogaea L.) for sugar profiling and oil quality. *Industrial Crops and Products, 51*, 46-50. https://doi.org/10.1016/j.indcrop.2013.08.050.

Chamberlin, K. D. (2019). Not your grandma's goobers: designing the future of peanut breeding. *Peanut Science, 46*(1A), 91-98.

da Costa Filho, P. A. (2009). Rapid determination of sucrose in chocolate mass using near infrared spectroscopy. *Analytica Chimica Acta, 631*(2), 206-211. https://doi.org/10.1016/j.aca.2008.10.049.

Damodaran, S., Parkin, K. L., & Fennema, O. R. (2007). *Fennema's food chemistry*: CRC press.

Deng, X., Liu, Z., Zhan, Y., Ni, K., Zhang, Y., Ma, W., Shao, S., Lv, X., Yuan, Y., & Rogers, K. M. (2020). Predictive geographical authentication of green tea with protected designation of origin using a random forest model. *Food Control, 107*. https://doi.org/10.1016/j.foodcont.2019.106807.

Fernandez-Novales, J., Garde-Cerdan, T., Tardaguila, J., Gutierrez-Gamboa, G., Perez-Alvarez, E. P., & Diago, M. P. (2019). Assessment of amino acids and total soluble solids in intact grape berries using contactless Vis and NIR spectroscopy during ripening. *Talanta, 199*, 244-253. https://doi.org/10.1016/j.talanta.2019.02.037.

Gardner, C. M. (2018). Transmission versus reflectance spectroscopy for quantitation. *Journal of Biomedical Optics, 23*(1), 1-8. https://doi.org/10.1117/1.JBO.23.1.018001.

Garrido-Varo, A., Sánchez-Bonilla, A., Maroto-Molina, F., Riccioli, C., & Pérez-Marín, D. (2018). Long-Length Fiber Optic Near-Infrared (NIR) Spectroscopy Probes for On-Line Quality Control of Processed Land Animal Proteins. *Applied Spectroscopy, 72*(8), 1170-1182. https://doi.org/10.1364/AS.72.001170.

Garvey, E. C., O'Sullivan, M. G., Kerry, J. P., Milner, L., Gallagher, E., & Kilcawley, K. N. (2021). Characterising the sensory quality and volatile aroma profile of clean-label sucrose reduced sponge cakes. *Food Chemistry, 342*, 128124. https://doi.org/10.1016/j.foodchem.2020.128124.

Gong, A., Shi, A., Liu, H., Yu, H., Liu, L., Lin, W., & Wang, Q. (2018). Relationship of chemical properties of different peanut varieties to peanut butter storage stability. *Journal of Integrative Agriculture, 17*(5), 1003-1010. https://doi.org/10.1016/s2095-3119(18)61919-7.

Hui, Y. H., Chen, F., Nollet, L. M., Guiné, R. P., Martín-Belloso, O., Mínguez-Mosquera, M. I., Paliyath, G., Pessoa, F. L., Le Quéré, J.-L., & Sidhu, J. S. (2010). *Handbook of fruit and vegetable flavors*. Hoboken: John Wiley and Sons.

Hussain, I., & Bowden, A. K. (2021). Smartphone-based optical spectroscopic platforms for biomedical applications: a review. *Biomedical Optics Express, 12*(4), 1974-1998.

Kundu, M., Khatkar, B. S., & Gulia, N. (2017). Assessment of chapatti quality of wheat varieties based on physicochemical, rheological and sensory traits. *Food Chemistry, 226*, 95-101. https://doi.org/10.1016/j.foodchem.2016.12.046.

Kusumaningrum, D., Lee, H., Lohumi, S., Mo, C., Kim, M. S., & Cho, B. K. (2018). Non-destructive technique for determining the viability of soybean (Glycine max) seeds using FT-NIR spectroscopy. *Journal of the Science of Food and Agriculture, 98*(5), 1734-1742. https://doi.org/10.1002/jsfa.8646.

Lan, W., Jaillais, B., Leca, A., Renard, C., & Bureau, S. (2020). A new application of NIR spectroscopy to describe and predict purees quality from the non-destructive apple measurements. *Food Chemistry, 310*, 125944. https://doi.org/10.1016/j.foodchem.2019.125944.

Liu, N., Parra, H. A., Pustjens, A., Hettinga, K., Mongondry, P., & van Ruth, S. M. (2018). Evaluation of portable near-infrared spectroscopy for organic milk authentication. *Talanta, 184*, 128-135. https://doi.org/10.1016/j.talanta.2018.02.097.

Mathlouthi, M., & Reiser, P. (2012). *Sucrose: Properties and Applications*. New York: Springer

Müller-Maatsch, J., Bertani, F. R., Mencattini, A., Gerardino, A., Martinelli, E., Weesepoel, Y., & van Ruth, S. (2021). The spectral treasure house of miniaturized instruments for food safety, quality and authen-

ticity applications: A perspective. *Trends in Food Science and Technology, 110*, 841-848. https://doi.org/10.1016/j.tifs.2021.01.091.

Müller-Maatsch, J., & van Ruth, S. M. (2021). Handheld Devices for Food Authentication and Their Applications: A Review. *Foods, 10*(12). https://doi.org/10.3390/foods10122901.

Nawade, B., Mishra, G. P., Radhakrishnan, T., Dodia, S. M., Ahmad, S., Kumar, A., Kumar, A., & Kundu, R. (2018). High oleic peanut breeding: Achievements, perspectives, and prospects. *Trends in Food Science & Technology, 78*, 107-119. https://doi.org/10.1016/j.tifs.2018.05.022.

Ortolan, F., & Steel, C. J. (2017). Protein characteristics that affect the quality of vital wheat gluten to be used in baking: A review. *Comprehensive reviews in food science and food safety, 16*(3), 369-381.

Phan-Thien, K.-Y., Wright, G. C., & Lee, N. A. (2010). Genotype-by-Environment Interaction Affects the Essential Mineral Composition of Peanut (Arachis hypogaea L.) Kernels. *Journal of Agricultural and Food Chemistry, 58*(16), 9204-9213. https://doi.org/10.1021/jf101332z.

Plazl, I., Leskovšek, S., & Koloini, T. (1995). Hydrolysis of sucrose by conventional and microwave heating in stirred tank reactor. *The Chemical Engineering Journal and the Biochemical Engineering Journal, 59*(3), 253-257.

Qi, H., Wu, Z., Zhang, L., Li, J., Zhou, J., Jun, Z., & Zhu, B. (2021). Monitoring of peanut leaves chlorophyll content based on drone-based multispectral image feature extraction. *Computers and Electronics in Agriculture, 187*. https://doi.org/10.1016/j.compag.2021.106292.

Rady, A., Fischer, J., Reeves, S., Logan, B., & Watson, N. J. (2019). The Effect of Light Intensity, Sensor Height, and Spectral Pre-Processing Methods when using NIR Spectroscopy to Identify Different Allergen-Containing Powdered Foods. *Sensors (Basel), 20*(1). https://doi.org/10.3390/s20010230.

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

Staniewski, B., Ogrodowska, D., Staniewska, K., & Kowalik, J. (2021). The effect of triacylglycerol and fatty acid composition on the rheological properties of butter. *International Dairy Journal, 114*. https://doi.org/10.1016/j.idairyj.2020.104913.

Tzounis, A., Katsoulas, N., Bartzanas, T., & Kittas, C. (2017). Internet of Things in agriculture, recent advances and future challenges. *Biosystems Engineering, 164*, 31-48.

Wang, L., Liu, H., Liu, L., Wang, Q., Li, S., & Li, Q. (2017). Prediction of peanut protein solubility based on the evaluation model established by supervised principal component regression. *Food Chemistry, 218*, 553-560. https://doi.org/10.1016/j.foodchem.2016.09.091.

Wang, Q. (2016). *Peanuts: processing technology and product development*: Academic Press.

Wang, Q. (2018). *Peanut Processing Suitability Evaluation Standards*.

Wang, Y., & Qi, H. (2015). *Food flavour chemistry*: Beijing Book Co. Inc.

Williams, P. (2021). Practical Aspects of Sampling for NIRS Analysis. In *Handbook of Near-Infrared Analysis* (pp. 375-414): CRC Press.

Workman, J., & Weyer, L. (2012). *Practical Guide and Spectral Atlas for Interpretive Near-Infrared*: CRC.

Xu, X., Xu, H., Xie, L., & Ying, Y. (2018). Effect of measurement position on prediction of apple soluble solids content (SSC) by an on-line near-infrared (NIR) system. *Journal of Food Measurement and Characterization, 13*(1), 506-512. https://doi.org/10.1007/s11694-018-9964-4.

Yan, J., Stuijvenberg, L., & 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). https://doi.org/10.1002/ejlt.201900031.

Yu, H., Liu, H., Erasmus, S. W., Zhao, S., Wang, Q., & van Ruth, S. M. (2020). Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics. *Industrial Crops and Products, 158*. https://doi.org/10.1016/j.indcrop.2020.112956.

Yu, H., Liu, H., Erasmus, S. W., Zhao, S., Wang, Q., & van Ruth, S. M. (2021). An explorative study on the relationships between the quality traits of peanut varieties and their peanut butters. *Lwt*. https://doi.org/10.1016/j.lwt.2021.112068.

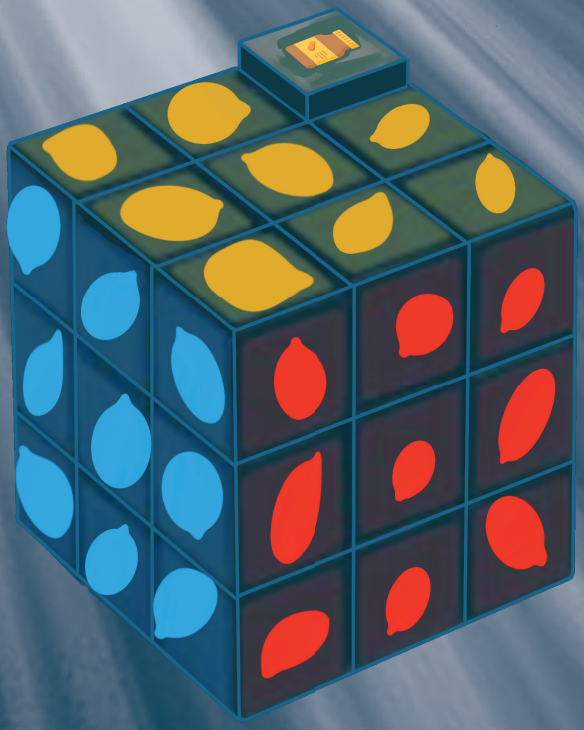
Yu, H., Liu, H., Wang, N., Yang, Y., Shi, A., Liu, L., Hu, H., Mzimbiri, R. I., & Wang, Q. (2016). Rapid and visual measurement of fat content in peanuts by using the hyperspectral imaging technique with chemometrics. *Analytical Methods, 8*(41), 7482-7492. https://doi.org/10.1039/c6ay02029a.

Yu, H., Liu, H., Wang, Q., & van Ruth, S. (2020). Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. *LWT - Food Science and Technology*, 109398. https://doi.org/10.1016/j.lwt.2020.109398.

Yuan, H., Bennett, R. S., Wang, N., & Chamberlin, K. D. (2019). Development of a Peanut Canopy Measurement System Using a Ground-Based LiDAR Sensor. *Front Plant Sci, 10*, 203. https://doi.org/10.3389/fpls.2019.00203.

Zhao, Z., Shi, A., Wang, Q., & Zhou, J. (2019). High Oleic Acid Peanut Oil and Extra Virgin Olive Oil Supplementation Attenuate Metabolic Syndrome in Rats by Modulating the Gut Microbiota. *Nutrients, 11*(12). https://doi.org/10.3390/nu11123005.

Zuo, Y., & Zhang, F. (2008). Effect of peanut mixed cropping with gramineous species on micronutrient concentrations and iron chlorosis of peanut plants grown in a calcareous soil. *Plant and Soil, 306*(1), 23-36.



SUMMARY

Peanuts as raw materials are the cornerstone of the whole peanut industry. The quality traits of peanuts determine their market values and the characteristics of final products, such as peanut butters. However, the quality traits of peanuts vary extensively with variety, growing environment, storage condition, and maturity. Meanwhile, the quality traits of peanuts have profound influences on the characteristics of their derived peanut butters. In order to understand the impact of the varieties on the traits of peanuts and peanut butters and develop rapid determinations, the main aim of this thesis is to elucidate and comprehend distinct analytical signatures and relationships of various types of peanuts and derived peanut butters and develop rapid methods of evaluation and identification of peanuts by near-infrared spectroscopy (NIRS) combined with chemometrics and machine learning.

In **Chapter 2** and **Chapter 3**, the different analytical signatures of batch samples and single peanut kernel peanuts were measured by conventional methods (e.g. gas chromatography and high-performance anion-exchange chromatography). The results showed that the fatty acids (FAs) composition of high oleic acid peanuts (HOP) and regular peanuts differed significantly. The models established by portable NIRS had the same performance to identify HOP and quantitatively measure the major FAs in batch peanut samples compared with benchtop NIRS. Meanwhile, considering the internal (measured by laser confocal microscopy) and external characteristics (kernel size) of different single peanut varieties, a single peanut detection accessory was designed and equipped with portable NIRS to collect spectral data of single peanut kernels. Based on the established quantitative models, breeding experts could quickly analyse the fat, protein, sucrose, and amino acids contents in single peanut kernels.

In **Chapter 4**, the relationships between peanuts and derived peanut butters were elucidated and the varieties were systematically clustered. It was found that lower fat and higher sucrose content in peanuts had great positive contributions to texture, rheology, and pyrazine content of peanut butters. Amino acids, especially serine, had positive effects on the main volatile compounds. Therefore, one group of peanuts (e.g. HY25, JH18, YH37, etc.) were applied to manufacture peanut butters with the highest pyrazine content and the highest values of texture and rheological properties. In **Chapter 5**, it was presented that peanut varieties were further clustered according to the structure characteristics (texture and rheology) and roast characteristics (colour and volatile compounds), respectively. The rapid identification methods of peanut varieties for peanut butter manufacture were established based on benchtop NIRS combined with machine learning. The sensitivity, specificity, and accuracy of cross validation and external validations using random forest and support vector machine algorithms were all over 90%, which offered new strategies for producers to rapidly identify peanut varieties for processing purposes.

In conclusion, the obstacles regarding the analytical fingerprints in batch and single kernel measurements as well as the relationship between peanuts and derived peanut butters, were addressed in this thesis. Furthermore, it is expected that the future development of novel spectroscopic approaches based on NIRS could boost critical links to improve the efficiency of peanut quality assessment. From breeding experts to producers, all stakeholders could have deep insights into the impacts of varieties on the quality traits of peanuts and their products. Ultimately, it will provide better control over the quality traits of peanut varieties in order to enhance quality, reduce waste, and mitigate integrity issues in the peanut (butter) supply chains.

ACKNOWLEDGEMENTS

Two thousand and forty-eight km was the number displayed on my phone to intendedly record the cycling distance during the COVID-19 pandemic. Cycling is probably one of the best sports against COVID-19 not only because of strengthening our body but also naturally keeping social distance and breathing in fresh air. Bikes have already become the main commuting tool since I was in high school. In addition to cycling around my hometown (Weihai, a cosy seaside city). The meaningful thing in high school was that I would be periodically responsible to manage classmates' bicycles. When I was in Beijing, with the development of shared bikes, cycling became an increasing efficient and sustainable travel option for me instead of the subway and cars. Sometimes I might cycle faster than the subway to the same destination whilst enjoying the scenery along the way. When I cycled in the Netherlands, the bike furtherly became an indispensable part of my life even in the unpredictable weather. When I wrote my propositions, I realised that I could cycle at a gallop alone but a relatively short distance. Conversely, I could cycle farther with teammate support and cooperation (e.g. from Wageningen to Den Hague, Rotterdam, or Dusseldorf). Just like the long ride, no one can finish the fantastic PhD trip without others' help. Now, it is an opportunity to thank those who contributed to my PhD establishment and supported me along my journey.

I would like to express my sincere gratitude to my WUR promotor Prof. **Saskia M. van Ruth** for your guidance, motivation, and immense knowledge. Without your approval, I would never have had this valuable opportunity to study at one of the best universities in Food Science. We had an excellent project discussion in person in China. Since then, from the first proposal draft to the final thesis, every discussion, comment, and email impelled me to think deeply and inspired me to be better researcher. By the way, I still owe you a cup of coffee for the thesis delivery.

I would like to offer my sincere gratitude to my CAAS promotor Prof. **Qiang Wang** for the continuous support of my PhD study and related research. Thanks for leading me on the path of scientific research since I started my master study. The contents of your speech at each public and private meeting are the treasure trove of prices containing the unique wisdom and philosophy of life and research, which is beneficial for my further life and work.

Besides my promotors, I would like to thank the rest of my thesis committee: Prof. **Henk A. Schols**, Dr **Jeroen Jansen**, Dr **Vincent Baeten**, and Dr **Di Wu** for their constructive comments and encouragement, but also for the insightful question which stimulated me to broaden my research from various perspectives.

I would like to extend my sincere gratitude to Prof. **Hongzhi Liu** for all the valuable scientific advice and continuous motivation. Working with you for almost ten years,

I imperceptibly felt good endeavour and the charm of your work and learned a lot from you. My gratitude is also to Dr **Sara W. Erasmus.** You are always altruistic, kind-hearted, and filled with unlimited positive energy. Without all of your vital input and help, the thesis would hardly be completed.

My Food Fraud buddies, **Alred, Valentina, Isabell, Ningjing, Yuzheng, Jing Yan, Zhijun, Lintianjiang, Qing Han, Jingjie, Laiyu, Pengfei, Wenyi, Anna, Luigi, and other partners**. Thanks for all the jokes, laughter, and support. I am glad to have your company and our friendship is 100% authentic.

I would like to thank all my colleagues in the peanut family: Dr **Yu Zhang**, Dr **Aimin**, Dr **Hui Hu**, Dr **Qin Guo**, Dr **Li Liu**, Dr **Bo Jiao,** Dr **Jinchuang**, Dr **Xiaojie**, Dr **Yu Song**, Dr **Shi Meng**, Dr **Weijing**, Dr **Kuijie**, Dr **Miao Yu**, Dr **Yan Chen**, Dr **Zuoyun**, Dr **Ning Li**, Dr **Zhihao**, Dr **Rui Guo**, Dr **Qiongling**, Dr **Liyou**, **Xiaolong, Nan Jiang, Ana, Fei Xu, Xiaojing, Lei Deng, Xiaoyong, Bingyu, Jiaxiao, Wenqiang, Qiuyan, Simeng, Genjiu, Yalong, Faisal, Manli, Xinyue, Tian Li, Yang Qu, Sisheng, Haodong, Yiying, Weiming, Yujie, Manzhu, Yue Liu, and other members.** Thank all of you for your constant support and assistance. Without you, life in Malianwa would have been too insipid. My gratitude is also to **Chuhong Li, Jinwu Gu, Jianxun Li, and other technicians** for your kindly help and interaction.

I also would to thank all my colleagues in the FQD group: **Alim, Ana, Anita, Annelies, Arianna, Ashkan, Ayusta, Bea, Christos, Chunyue, Cristina, Cuirong, Diana, Ebru, Edoardo, Erik, Eva, Fabiola, Feilong, Femke, Fiametta, Frans, Folake, Hannah, Hao Zhang, Ita, James, Jilu, Jianing, Jiaying, Jonna, Julie, Kangni, Kasper, Keqing, Li Wang, Lijiao, Ling Xiong, Luc, Lucia, Lu Ying, Marianna, Maryia, Matthijs, Melania, Moheb, Mostafa, Onu, Peiheng, Pieter Dekker, Pieter Groen, Qing Ren, Ranti, Ruth Boachie, Ruth Ngadze, Rutger, Ruud, Sam, Sara Pacheco, Sergio, Shiksha, Sofia, Sine, Swantje, Sydney, Teresa, Xiangnan, Yajing, Yifan, Yao Chen, Yongkai, Zekun, Zhan Huang, Zhe Cheng, Zongyao, and other FQD peeps.** Our group is international and diverse. Every talk was a cultural exchange. Every joke was an emotional empathy. Every sports activity (e.g. WeDay or table tennis) was the release of stress. Every trip (e.g. PhD trip to Australia) was a positive interpersonal interaction. Thank you to all for filling my road with rainbows and sweetmeats. My gratitude is also to **Corine, Kimberley,** and **Lisanne.** You are brilliant, versatile, and warm-hearted. Any trivia is a piece of cake for you. I really appreciate your help. Special gratitude is dedicated to Prof. **Vincenzo** for your leading such a perfect group and the tricks from email, WhatsApp, and LinkedIn.

I would like to thank **Mingjun Zhang** for the efforts from your side. Every conversation with you inspires and incents me. My gratitude is also to **Claudius van de Vijver.**

Thank you for the impressive introduction about what PhD is and bicycle storing service. Cycling with you was a pleasant experience.

I would like to thank **Yvonne, Vesna,** and the other VLAG colleagues, **Yiwei Dong, Qiongyue Zhang, Xiaochen Zhang, Xinmiao Liu,** as well as the other CAAS colleagues for contributing to my project. Without your help, the project would be stuck.

I would like to thank my other buddies from the CAAS-WUR project: **Siyuan, Yifan, Chen Zhang, Xiulu, Qian Chen, Tian Lin, Fang Wang, Xiaoqian, Hongrui, Qiuhong, Fanrong, Fen Xu, Jiaqi Huang, Jing Jin, Bo zhou, Xue zhang, Chenlong, Tianpeng, Hao Chen, Jiaqi Zhang, Dazhi, Lina**, etc. Thanks for all the fun, laughter, inspiring discussions, and various activities. You made my PhD life easier and happier.

I would like to thank my roommates (Fengzeyinghe 17-603, Asserpark 44-02C, Haarweg 11, and Dijkgraaf 4-09C) for our best dwelling condition, cultural exchange, Dutch learning, and food tasting, especially **Lucas, Thijs, Wester, Tom, Laura, Gui, Rhianne, Vita, Joann, Ine, Bent, Ojasvi, Qifa, Xiaoqi, Leon**, and **Holger** (a cute hamster). Our corridor 09C is the best in Dijkgraaf 4.

I would like to thank my paranymphs: **Mohammad Almansouri** and **Liang Fang**. **Momo**, you are my superman. When my ride was stuck in Arnhem, you came immediately to help me. I learned a lot from you about the food, language, and the culture of Saudi Arabia and Muslims. **Liang**, you are my Michelin chef, chief stylist and health coach. Without you, life in Wageningen would be mundane. Thanks for sharing about cuisine, coffee, and liquor.

I would like to thank my sports buddies, **Shenyi, Yaowei, Jian Liu, Hui Liu, Qingyue, Huayi,** and **Yang Shen**. You were my support and motivation during the COVID-19 pandemic. The discussion and action about the balance between delicacies and sports always exist but sensibility would always be greater than reason.

I would like to thank the team members of the IFST basketball club (**Xuanming Tang**, **Zhengguo Yong**, **Haidong Ding**, etc.) and the IPO basketball team (**Li Xie, Kees, Anne-marie, Leonore, Violette, Herman, Huiju, Ke Meng, Yujie, Yunyan, Nikola,** etc.). Every Wednesday or Thursday evening was a cosy night for me to have deep relaxation. Many thanks for making the shot in my assist and vice versa.

I would like to thank excellent friends in China and the Netherlands, **Wei Xu, Xilong, Taojun, Libin, Haibo, Dan Kirk, Ruihao, Nina, Xiyu, Fengzheng, Wenjie, Xiangyu,**

Site, Ruifang, Wenbo, Han Zhang, **Weiqing**, **Pengfei Zhang, Bo Shang, Haiyang, Liang Wu, Yang zhao, Bin Tian, Xiao Fei, Junya, Yang Zhang, Yuan ye, Yecheng, Shicai, Yuanqing, Lei Ma, Jianju, and other friends**. Thank you all for adding beautiful scenery to my journey.

I would like to thank **Hongri Sun, Zuojiang Yu, Guosheng Zhang**, and other aunts and uncles. You are not only my respected elders but also my mentors and friends.

I would like to thank my beautiful wife, Mrs. **Ju Dong**. Marrying you is as natural as breathing. Thanks for your selfless dedication. You were the light of hope when I was in dark. You were the sound nature when I was down. I love you!

The whole life story began all because two people fell in love. I would like to thank my **parents** and **parents-in-law** for their unconditional support and love. The happiness of growing up is my shield and home are always my safe harbour. I would like to thank all my **relatives**. All of you are my strong backing.

There are still many people not included in this acknowledgement. I apologise for that. Please trust that my appreciation for you is no less than those listed above.

"The Last Question", is a science fiction short story written by Isaac Asimov. The last question is "Can entropy ever be reversed?". For me, I would like to ask myself the last question as to the end. Can I cycle ever reversed?

Axis Y1010, Wageningen University
Hongwei

ABOUT THE AUTHOR

CURRICULUM VITAE

Hongwei Yu was born in Weihai, Shandong Province, China, on the 1st February 1991. He obtained his bachelor degree in Food Science and Engineering at Qingdao Agricultural University in 2013. After three years study at the Chinese Academy of Agricultural Sciences (CAAS), he obtained his Master degree in Food science. In February 2017, he started his Sandwich Doctoral studies on the novel spectroscopic approaches for the characterisation of quality- and identity-related key features of peanuts and peanut butters at Wageningen University and Research (WUR) and CAAS, under the supervision of Prof. Dr Saskia M. van Ruth from WUR and Prof. Dr Qiang Wang from CAAS. The results of his PhD project are described in this thesis. His main research interests cover topics in food quality and data analysis.

LIST OF PUBLICATIONS

This thesis:

Yu, H., Liu, H., Wang, Q., & van Ruth, S. M. (2020). Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. LWT - Food Science and Technology, 128, 109398.

Yu, H., Liu, H., Erasmus, S. W., Zhao S., Wang, Q., & van Ruth, S. M. (2020). Rapid high-throughput determination of major components and amino acids in a single peanut kernel based on portable near-infrared spectroscopy combined with chemometrics. Industrial Crops and Products, 158, 112956.

Yu, H., Liu, H., Erasmus, S. W., Zhao S., Wang, Q., & van Ruth, S. M. (2021). An explorative study on the relationships between the quality traits of peanut varieties and their peanut butters. LWT - Food Science and Technology, 151, 112068.

Yu, H., Liu, H., Erasmus, S. W., Wang, Q., & van Ruth, S. M. Rapid classification of peanut varieties for its processing into peanut butters based on near-infrared spectroscopy combined with machine learning. Submitted.

Other publications:

Gong, A., Shi, A., Liu, H., Yu, H., Liu, L., Lin, W., & Wang, Q. (2018). Relationship of chemical properties of different peanut varieties to peanut butter storage stability. Journal of integrative agriculture, 17: 1003-1010.

Wang, Q., Liu, H., Shi, A., Hu, H., Liu, L., Wang, L., & Yu, H. (2017). Review on the processing characteristics of cereals and oilseeds and their processing suitability evaluation technology. Journal of integrative agriculture, 16: 2886-2897.

OVERVIEW OF COMPLETED TRAINING ACTIVITIES

Discipline specific activities

- The Unexpected Science of Chocolate and Steel (2017). Delft, The Netherlands.
- Smartphone-based Food Analysis (2017). Wageningen, The Netherlands.
- Metrology in Food and Nutrition (2017). Prague, Czech Republic.
- China Peanut Industry Development Conference (2018). Qingdao, China. Oral presentation.
- Food Integrity Training Week: Symposium for Postgraduate Students on Food Fraud (2018). Prague, Czech Republic. Poster presentation.
- Ninth Annual Academic Conference of CCOA (2018). Beijing, China.
- Food Authentication and Guarding Against Unknown Adulterants (2019). Online.
- Philosophy and Ethics of Food Science & Technology (2020). Wageningen, The Netherlands.
- Advanced Organic Chemistry (2020). Online.
- The VLAG Online Lecture Series (2020). Online.
- Workshop Smart Tech for Food (ST4F) (2020). Online.
- OLEUM H2020 Final Conference (2021). Online.
- Recent Advances in Food Analysis (2021). Online.

General courses

- Academic writing and presenting in English (2016). Beijing, China.
- Efficient and Effective Academic Development Beijing (2016). Beijing, China.
- WGS PhD workshop Carousel (2017). Wageningen, The Netherlands.
- Scientific/Proposal Writing (2017). Wageningen, The Netherlands.
- Statistical Programming with R (2017). Utrecht, The Netherlands.
- RMarkdown (2020). Wageningen, The Netherlands.
- Bioinformatics with Linux and Python (2021). Online.
- Big Data in the Life Sciences (2021). Wageningen, The Netherlands.

Optional courses and activities

- Preparation of research proposal (2017). FQD, Wageningen, The Netherlands.
- Weekly group meetings/Project meetings with stakeholders (2017-2021). FQD, Wageningen, The Netherlands and CAAS, Beijing, China.
- Colloquia (2017-2021). FQD, Wageningen, The Netherlands.
- FQD PhD trip to Australia (2018). FQD, Wageningen, The Netherlands.
- Food Fraud and Mitigation (2020). Online.

List of abbreviations

CAAS: Chinese Academy of Agricultural Sciences
CCOA: Chinese Cereals and Oils Association
FQD: Food Quality and Design
VLAG: Graduate school for Nutrition, Food Technology, Agrobiotechnology, and Health Sciences
WGS: Wageningen Graduate School
WUR: Wageningen University & Research

The research described in this thesis was financially supported by the National peanut industry technology system “Research and application of key technologies for peanut processing suitability evaluation, quality improvement and efficiency increase” (CARS-13), Hebei oil innovation team of modern agro-industry technology research system (HBCT2019090203), Weihai Wendeng District Industry-University-Research Distinguished Expert Support Program “Evaluation of the suitability of peanut butter and high-quality product development and industrialization demonstration”, CAAS-WUR Joint PhD Programme, and the China Scholarship Council (CSC).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Hongwei
Printed by ProefschriftMaken