Visible Hyperspectral Imaging for Predicting Intra-muscular Fat Content from Sheep Carcasses

by

Joseph Aduol Sigar

Supervisor: Dr Nicola Armstrong

A thesis submitted for the degree of Honours in Mathematics & Statistics.

February 10, 2020

SCHOOL OF ENGINEERING & INFORMATION TECHNOLOGY
MATHEMATICS & STATISTICS

Murdoch UNIVERSITY

Contents

Abstract

Intramuscular fat (IMF) content plays a key role in the quality attributes of meat, such as sensory properties and health considerations. The tenderness, flavour and juiciness of meat are examples of sensory attributes influenced by IMF content. Traditionally, IMF content in meat was determined using destructive, time consuming and at times unsuitable methods in industry applications. However, with recent advancement of technology, there has been an interest in exlporing ways to ascertain meat quality without damage. Hyperspectral imaging analysis is an emerging technology that combines the use of spectroscopy and computer imaging analysis to obtain both the spectral and spatial information of objects of interest. Hyperspectral imaging was initially developed for remote sensing, but has recently emerged as powerful tool for non-destructive analysis of quality in the food industry and has had very accurate results in the prediction of meat qualities such as IMF content. In this thesis, we use a data set of 101 hyperspectral images of sheep carcasses to investigate the ability of multivariate statistical methods to accurately predict IMF content.

Signed Statement

This work contains no material which has been accepted for the award of any other degree or diploma in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Signed: ..... Date: .....

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisor Dr. Nicola Armstrong of the School of Engineering and Information Technology at Murdoch University for the continuous support of my Honours degree study and thesis research, for her patience, motivation and vast knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Honours thesis.

Besides my supervisor, I would like to thank the rest of my thesis committee: Mr Jeremiah Sigar and Mr Shelton Mudzi, for their insightful comments and encouragement. A special thanks to Walter Sigar and Rose Aduol for generously offering their time and skills on short notice.

Last but not least, I would like to thank my family, partner and friends, for supporting me throughtout writing this thesis and my life in general.

Chapter 1

Introduction

Hyperspectral Imaging is an emerging technique that combines conventional spectroscopy imaging and computer imaging techniques, enabling an in-depth analysis of images and objects that neither of the techniques can achieve independently. hyperspectral imaging has also been hailed as a non-invasive, non-destructive and efficient technique in various fields such as medicine (Lu and Fei (2014); Leavesley et al. (2016); Halicek et al. (2018)), pharmaceutics (Ravn, Skibsted, and Bro (2008); Roggo et al. (2005)) and agriculture (Mehl et al. (2004); Gowen et al. (2007); C.-H. Feng et al. (2018)). In comparison, traditional techniques such as human inspection or high performance liquid chromatography, are deemed destructive, inconsistent and time-consuming (D. Wu and Sun (2013a)).

Hyperspectral imaging is a branch of the spectral imaging family that collects and analyses stacks of images. It was first used in the identification and classification of various minerals such as carbonate minerals or iron oxides in remote sensing (Cloutis (1996); Sabins (1999)). Other applications in remote sensing have been the study of planetary objects (Van der Meer et al. (2012); C. A. Lee et al. (2011)) and to correctly identify and map wetland vegetation (Adam, Mutanga, and Rugege (2010)).

With advancements in computational processing power and technology (both hardware and software), hyperspectral imaging became popular in other fields. For example, it has been employed in pharmaceutical research and production, to help identify and classify chemical components objects such as mapping cutaneous tissue oxygen concentration (Miclos et al. (2015)). Pavurala, Xu, and Krishnaiah (2017) used hyperspectral imaging to monitor the uniformity of coat thickness in the manufacture of transdermal drugs, while a study published by Wilczyński et al. (2016) built a hyperspectral imaging tool to detect and identify counterfeit drugs.

In the medical discipline, Calin et al. (2015) and Wahabzada et al. (2017) were able to characterize and monitor wounds in patients using hyperspectral imaging. Leavesley et al. (2016) developed a hyperspectral imaging fluorescence excitation scan that enabled the classification and detection of colon cancer in patients. Halicek et al. (2018) used hyperspectral imaging with neural networks to classify tumor margins in head and neck cancer. And recently, I.-C. Wu et al. (2018) were able to use hyperspectral imaging in the early identification of esophageal squamous neoplasm (esophageal cancer).

In agriculture, hyperspectral imaging has replaced traditional techniques , which were found to be destructive, tedious, reliant on human intervention, time-consuming and sometimes environmentally unfriendly (Huang, Liu, and Ngadi (2014)). Hyperspectral imaging offers various advantages compared to traditional techniques, with the most significant advantage being its prediction accuracy. For instance, the quality and physical attributes of fruits and vegetables can be identified and predicted or classified with greater accuracy using hyperspectral imaging (Nicolai et al. (2007); Oliveira and Franca (2011)). In the meat industry, Kamruzzaman, Makino, and Oshita (2016) were able to successfully predict the moisture content in meat, while Lohumi et al. (2016) developed a calibrated model that could predict meat fat content. Naganathan et al. (2016) built a forecasting model that predicted the tenderness of beef using hyperspectral imaging. These examples of the accuracy of hyperspectral imaging and others, have allowed the meat processing industry to build tools to rapidly classify quality levels of meat, reducing object detecting time, saving both cost and time on potentially destructive laboratory techniques (C.-H. Feng et al. (2018))

There are numerous advantages to the application of hyperspectral imaging as seen above. The aim of this chapter is to give an overview of what hyperspectral imaging analysis is, to explore some of the equipment used for hyperspectral imaging applications and finally to provide an overview of the standard analysis framework performed for acquired hyperspectral data. This chapter is divided in the following way: The first section (1.1) gives a basic overview of the techniques that hyperspectral imaging is derived from, then alternatives are introduced to hyperspectral imaging such as multispectral and full-spectral imaging. Exploration of the various hyperspectral image acquisition modes and the underlying data structures of a hyperspectral image is also introduced.

As hyperspectral images cannot be taken without special equipment, the next section (1.2) gives a description of the various components used to acquire hyperspectral images. And finally, the last section (1.3) focuses on the anal-

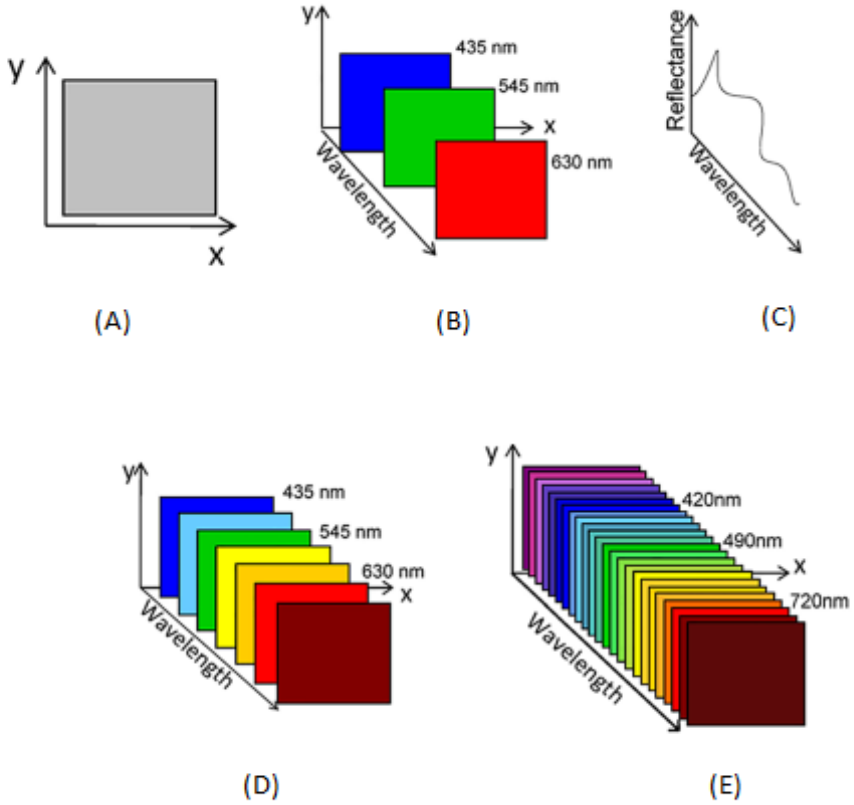


Figure 1.1: Different imaging methods. (A) A two-dimensional image of a sample showing the spatial (x,y) distribution. (B) RGB colour image with only three bands (red, green and blue wavelengths respectively). (C) Spectroscopy graph showing the spectral distribution. (D) A multispectral three-dimensional hypercube showing both the spatial and spectral distribution on a limited number of spectral bands. (E) A hyperspectral hypercube with a contiguous spectrum. Figure adapted from Li et al. (2013)

ysis framework for hyperspectral images. This section covers pre-processing techniques required prior to analysis such reflectance calibration, spectral and spatial pre-processing steps and image processing.

1.1 What is Hyperspectral Imaging?

To get an understanding about the underlying structure of hyperspectral imaging, a brief introduction to the two fundamental techniques that it is derived from is required. These techniques are computer vision and spectroscopy.

Computer vision, sometimes referred to as imaging analysis or computer imaging, is a technique that operates by analysing an image on the visible wavelength, particularly the red (564–580nm), blue (534–545nm) and green (420–440nm) wavelengths. By operating on the visible wavelength, computer

vision is able to extract shape-based features of objects within the image and perform analysis of the image (D. Wu and Sun (2013b), Abdullah et al. (2004)). Recently, computer vision has been employed in the automotive industry to build self-driving cars that rely partly on computer vision to display information around the car. This information is then analysed and fed into the car's artificial autonomous system for further action, such as obstacle detection and avoidance (Chen et al. (2015); Hadsell et al. (2009)).

Fig. 1.1 (B), shows the wavelengths that a standard computer vision system can be applied to. However, computer vision has some big drawbacks including the inability to accurately classifying two or more objects that have roughly the same physical characteristics, such as the same shape and/or colour. It is also unable to detect invisible defects on objects within an image or detect chemical components of objects for further analysis. To overcome these limitations, computer vision is often coupled with other techniques in order to increase accuracy (C. Zheng, Sun, and L. Zheng (2006); D. Wu and Sun (2013b); Du and Sun (2006)).

Spectroscopy is a technique that measures the interaction between electromagnetic radiation (simply known as optical properties) and light in the visible wavelength range (340–740nm) to determine the chemical and physical characteristics present in objects of interest (Bock and Connelly (2008); Cen and Y. He (2007)). It categorizes an object's interaction with the visible light range into either transparent, translucent or opaque objects. In this context, transparent indicates an object with little or no absorption and reflection, while opaque objects tend to have high absorption and reflection rate. Translucent objects have an absorption and reflection rate that is roughly between transparent and opaque (Amigo, Babamoradi, and Elcoroaristizabal (2015)). Recent studies have also considered using the near infra-red wavelength range (750–2500nm) instead of the visible wavelength range, and have found promising results in determining the chemical and physical characteristics in objects of interest (Ravn, Skibsted, and Bro (2008); Oliveira and Franca (2011)).

Fig. 1.1 (C) shows a reflectance against wavelength plot, that measures how much light has been reflected by an object of interest. This is produced in such a way that as light photons enter the object, some are reflected (reflectance), some pass through (transmission), some are refracted (refraction) and some are absorbed (absorption) (Gowen et al. (2007)). This process identifies the optical properties that are measured and analysed to derive information about the chemical characteristics of an object from the reflected light (Cen and Y. He (2007)). However, while computer vision is able to show the spatial distributions of an object within an image at selected wavelength bands (Fig. 1.1

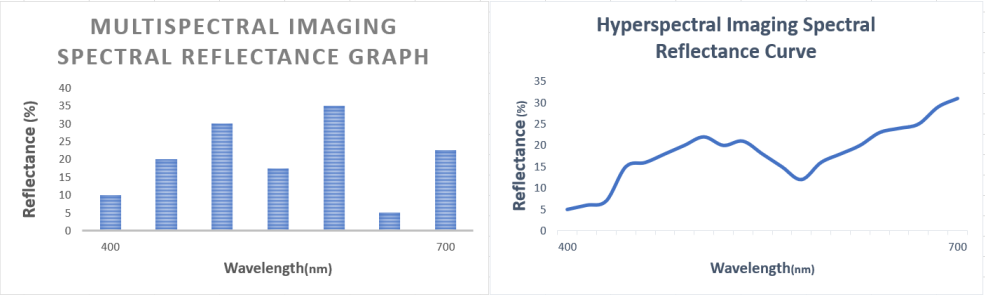


Figure 1.2: Spectral Reflectance Graph/Curve of a multi-spectral image (left) and that of a hyperspectral image (right).

A), spectroscopy is incapable of displaying the spatial distribution of an object. This limits where the technique can be applied.

Hyperspectral imaging integrates the advantages of computer vision and spectroscopy into one system, whereby it can extract both the spatial and spectral distributions from an image, acquired using a camera with a spectrometer attached to it. That is, hyperspectral imaging is a method that acquires an image in various spectral (wavelength) bands that provides information on the inherent chemical and physical attributes of the object of interest. When coupled with multivariate analysis methods, hyperspectral imaging is capable of accurately displaying the content distribution of the objects within an image as well as providing accurate classification and prediction of objects in the image (Gowen et al. (2007); Kamruzzaman, G. ElMasry, et al. (2012a)).

1.1.1 Principles of Hyperspectral Imaging

Types of spectral imaging

Types of spectral imaging include multispectral and full-spectral (also referred to as ultra-spectral) imaging (Gowen et al. (2007); D. Wu and Sun (2013a)). These methods share the same underlying concept as hyperspectral imaging with the difference being the number of spectral bands available in the spectral image.

In multispectral imaging, Fig 1.1 (D), the number of spectral bands is normally ten or less and in most cases cannot provide a real spectrum for every image. As such, the spectral resolution of a multispectral imaging system tends to be 10 nanometers (nm) or larger, as illustrated by the left plot in Fig 1.2. Spectral resolution measures the ability to resolve spectral features and bands into their separate components. A resolution that is too low results in lost spectral information, preventing correct identification and characterization of the object, while a high resolution may result in a longer measurement time. The

Table 1.1: Differences among RGB Imaging, Spectroscopy, Multispectral and Hyperspectral Imaging techniques

Features	Computer Imaging	Spectroscopy Imaging	Multispectral Imaging	Hyperspectral Imaging
Spectral information	×	✓	Limited	✓
Number of Spectral bands	3	Several to hundreds	3 to 10	Several to hundreds
Spatial information	✓	×	✓	✓
Multi-constituent information	×	✓	Limited	✓
Detection of small objects	×	✓	Limited	✓

choice between low or high resolution is dependent on the application and the desired information from the experiment (Huang, Liu, and Ngadi (2014); Kamruzzaman, G. ElMasry, et al. (2012b); Van der Meer et al. (2012)).

Hyperspectral imaging systems tend to acquire hundreds of images of the same object, with contiguous/connecting wavelengths (with a spectral resolution of less than 10nm), as shown in Fig 1.1 (E). Every pixel in the hyperspectral image contains its own spectrum over a contiguous wavelength range, as illustrated by the reflectance graph on the right of Fig 1.2. This improves the accuracy of prediction/classification produced by a hyperspectral analysis as opposed to multispectral analysis, but at the cost of image acquisition speed. The ultra-spectral (full-spectral) imaging system is the successor of hyperspectral imaging systems, with an even finer spectral resolution, allowing an even finer representation over the contiguous wavelength range (D. Wu and Sun (2013a)).

Table 1.1 summarizes the differences between computer imaging, spectroscopy, multispectral imaging and hyperspectral imaging, as discussed above, highlighting the difference between multispectral and hyperspectral imaging.

Hyperspectral Data

The spectral cube (also known as hypercube, data-cube or hyperspectral cube) is a three-dimensional (3-D, (x, y, z)) block of data that is comprised of a two-al (2-D, (x, y)) image of the object and the spectral information on the 3^{rd}-dimension (usually denoted by λ). The spectral cube is formed by obtaining the 2-D image (spatial information) of the different spectral bands one at a time, and then the 3-D image is formed by stacking the 2-D images together. This is illustrated in Fig. 1.1 (D) and (E). Note that for the remainder of this thesis, we will use the terms spectral bands and wavelengths interchangeably.

Image acquisition

To acquire the 3-D hyperspectral image cube, there are 4 main approaches commonly used in research and industry. These approaches are the whisk-broom method, push-broom method, area scanning and single shot method (Lu and Fei (2014); Gowen et al. (2007); Adão et al. (2017)). The whisk-broom method, commonly referred to as point scanning, allows for an individual point to be scanned, providing the required spectrum of this point. Then either the camera is moved to another point along the spatial dimension of the object or the object is moved to allow the camera to capture the spectrum at another point. This process is repeated until the entire object has been scanned. An illustration of the whisk-broom method is depicted in Fig. 1.3 (A).

Push-broom scanning (line scanning), scans a whole line of the object along one spatial dimension (either x or y) capturing the spectral information corresponding to each pixel on the line (Fig. 1.3 (B)) . To construct the hypercube in a push-broom scanning approach, the object is then moved in a direction orthogonal to the scanned line.

The area scanning approach differs from point and line scanning, because instead of having to move either the object or the camera, it acquires a monochrome 2-D spatial image of the object at a single wavelength at a time. The 3-D hypercube is then constructed by repeating the scanning over the whole wavelength range (Zuzak et al. (2001)). The single shot mode is a fairly new method, that has been shown to be capable of capturing both the spatial and spectral information using a large area detector to capture the images. However, it is still in the early stages of development, and it currently is limited in acquiring spatial and spectral information (D. Wu and Sun (2013a), Adão et al. (2017)).

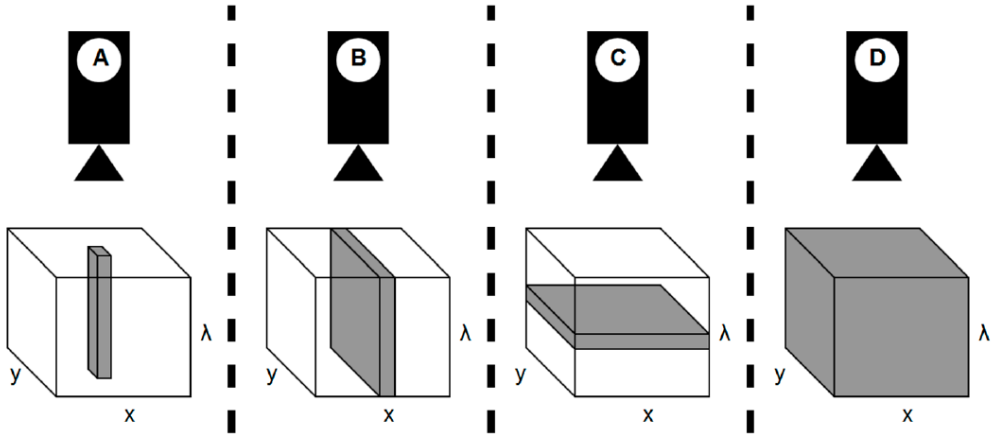


Figure 1.3: Hyperspectral Imaging Acquisition modes. (A) represents point scanning or whisk-broom method; (B) represents line scanning or push-broom method; (C) represents area scanning and (D) represents single shot, a fairly new approach. Adapted from Adão et al. (2017)

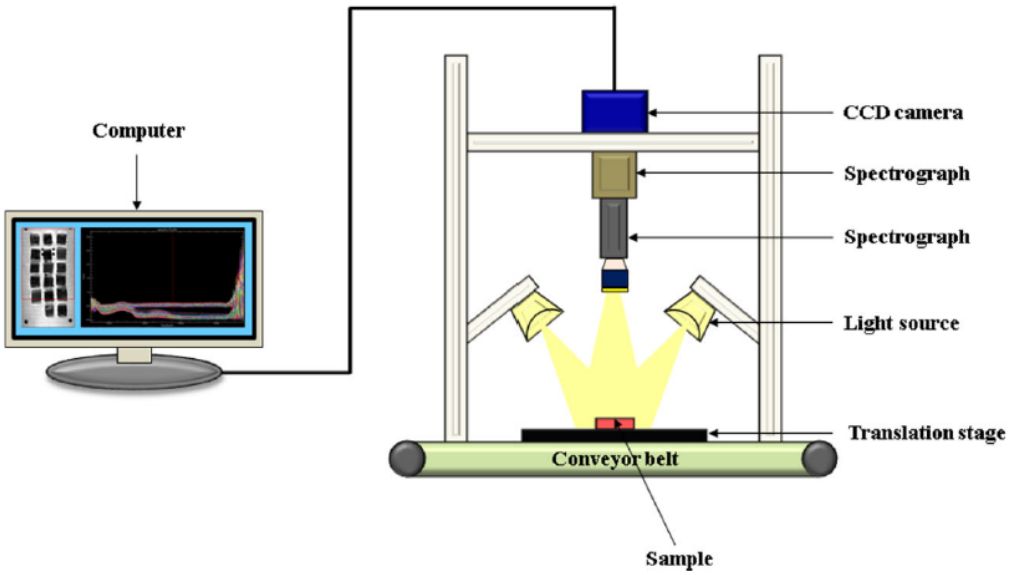


Figure 1.4: Schematic of the Hyperspectral Imaging configuration. Adapted from H.-J. He and Sun (2015)

1.2 Hyperspectral Imaging Equipment

The schematic of hyperspectral imaging equipment is illustrated in Fig. 1.4. The main components include a computer with software to control the image acquisition process, a light source, a spectrograph, a charge-couple device camera and a translation stage.

The light source generates the required light that acts as an information carrier to illuminate the target. It is an integral part of the hyperspectral sys-

tem, as it allows the detection of the physical structure and the chemical components of the object of interest (Nanyam et al. (2012); Y.-Y. Pu, Y.-Z. Feng, and Sun (2015); H. Pu, Kamruzzaman, and Sun (2015)). Various light sources are used depending on the application of the hyperspectral technique such as halogen lamps, lasers, tunable light and light emitting diodes. D. Wu and Sun (2013a) examined each of these sources and their advantages and applications in detail.

The camera used in the image acquisition stage has the ability to quantify the intensity of the captured light. Quantifying is done by converting the light photons into electrons. Two widely used cameras in hyperspectral imaging are the charge-coupled device (CCD) and the complementary metal oxide-semiconductor (CMOS) (D. Wu and Sun (2013a), C.-H. Feng et al. (2018), Huang, Liu, and Ngadi (2014), Gowen et al. (2007)). A lens may be included in the imaging equipment to ensure there is enough focus as well as to potentially delineate the field of view (C.-H. Feng et al. (2018); Kamruzzaman, Makino, and Oshita (2015)). The spectrograph is another important component used in the image acquisition stage. It is mainly used to disperse captured light into different continous spectral ranges. The captured light is usually detected by a charge-couple device camera. Using the camera in conjunction with the spectrograph, as stated by both Kamruzzaman, Makino, and Oshita (2015) and C.-H. Feng et al. (2018), allows a limited amount of light to reach the camera, which is controlled by the spectrograph. D. Wu and Sun (2013a) state that a spectrograph is one example that can be used in dispersing the light captured. Other commonly used examples include filter wheels, acousto-optic tunable filters, liquid crystal tunable filters, single shot imagers and Fourier transform imaging spectrometers.

Depending on the acquisition mode, the translation stage is movable to allow the spatial dimension to be formed with the object moving. Finally, the computer with software to control the image processing, is used when further analysis is required. It can also be used to control the exposure time and wavelength range, store any acquired images and develop models from the stored data. (C.-H. Feng et al. (2018)).

1.3 Analysis framework

The process involved in hyperspectral data analysis differs greatly depending on the objects and/or characteristics of the object of interest. This means that a highly accurate process applied to a hypercube of object A would not yield the same level of accuracy when applied to another object B. As such, several

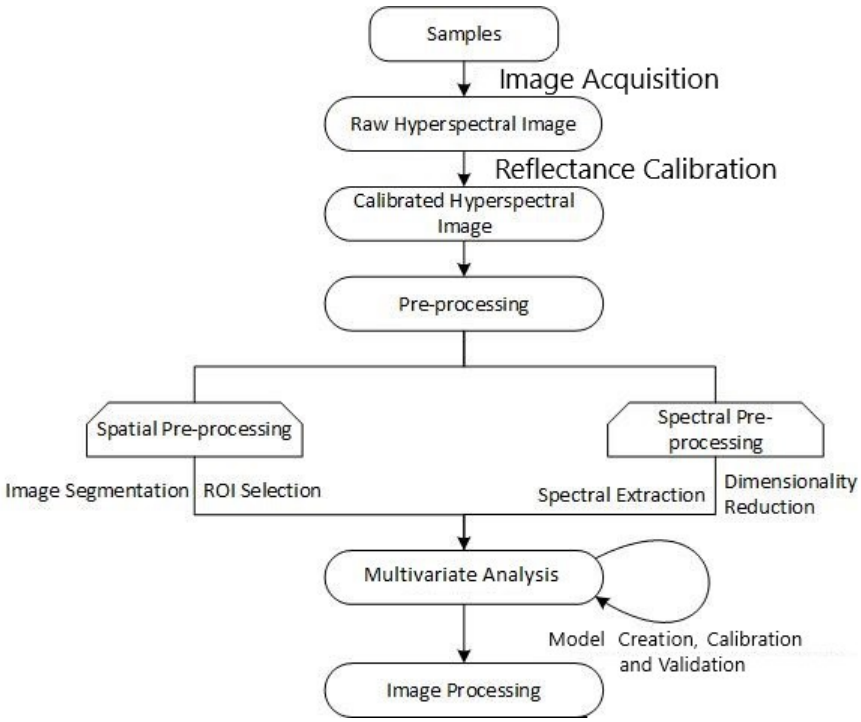


Figure 1.5: Flowchart of typical steps used when analyzing Hyperspectral Imaging Data

studies have proposed basic or standard guidelines to be used when analysing hyperspectral images (Amigo, Babamoradi, and Elcoroaristizabal (2015); G. M. ElMasry and Nakauchi (2016); Kamruzzaman, G. ElMasry, et al. (2012b); D. Wu and Sun (2013a); Huang, Liu, and Ngadi (2014); Gowen et al. (2007)). These are shown graphically in Figure 1.5. Initially a raw hyperspectral image (hypercube) is obtained using one of the image acquisition techniques discussed in section 1.1.1. The raw hypercube is then calibrated using reflectance calibration to minimize any camera effect. The calibrated hypercube then undergoes spatial and spectral pre-processing that allows the selection of regions of interests (ROIs) and spectral extraction. After this, a qualitative or quantitative model is created, calibrated and validated using a multivariate statistical technique. This section provides an overview of these steps.

1.3.1 Reflectance Calibration

Reflectance calibration converts the raw intensity value of the raw hyperspectral image obtained by the camera sensor to relative reflective values. This correction minimizes the effect of the residual electric current flowing in a camera (dark current), as well as minimizing the influence of the background spectral response of the spectrography (G. M. ElMasry and Nakauchi (2016)).

1.3.2 Pre-processing

Spatial Pre-processing (Image Segmentation)

Spatial pre-processing, commonly referred to as image segmentation, is a very important step in any image analysis procedure, since it has been found that any subsequent analysis is dependent on this operation (Sun and Du (2004)). This step allows the extraction of object/s of interest from an image, for which further analysis is to be carried out. It also allows for edges of other objects in the image to be detected, as well as the use of techniques that would improve the image quality of the objects of interest (G. ElMasry, Wang, and Vigneault (2009)). It is usually at this stage of a hyperspectral data analysis that Region of Interests (ROIs) are selected. That is, the regions which the researchers are interested in analysing in detail (Kamruzzaman, G. ElMasry, et al. (2013)). The general basic procedure of selecting a ROI involves creating a mask with the same spatial dimension as the image, however only information for the pixels in the ROI are preserved. This can vary from creating a matrix of *0's* and *1's*, where *1's* represent the ROI, to advanced techniques such as partial-differential equation based methods, edge detection, histogram-based methods, Fourier transforms (FT), wavelet transforms (WT) or thresholding (Esquerre et al. (2012); Bong and Rajeswari (2011); Zhang (1996)).

Of the aforementioned advanced techniques, thresholding is commonly used in image segmentation due to the simplicity of implementation. It is achieved by selecting a threshold value that allows a clear contrast between the object/s of interest and the background. This is done by denoting pixels with intensities above a threshold value as *1* and the remaining pixels are set to *0*, indicating background. However, selection of the threshold value can require human intervention and can be a challenging, time-consuming and tedious task (Yang et al. (2007); Samopa and Asano (2009)). This is because there is a compromise in the selection of a threshold value, between choosing a low value to increase the possibility of distinguishing the main object from the background and choosing a high value to remove the background. The process is repeated for each wavelength in the hypercube. It is prone to error if there are many bands in the hypercube (with each band requiring a different threshold value). To make the selection of the threshold value automated with no human intervention, an automatic threshold scheme can be used. Such a scheme allows for image detection and the selection of ROIs by setting an initial threshold value based on their gray-level distribution (Al-Amri, Kalyankar, et al. (2010); Ng (2006)).

Spectral Pre-processing

Spectral pre-processing is used to improve the extracted spectral data by both avoiding the influence of undesirable effects coming from the measurement or instruments, such as light scattering or spectral noise, and distinguishing the differences between the spectra that would help in any subsequent analysis. Methods that can be applied during spectral pre-processing include smoothing, derivatives, standard normal variates (SNV), multiplicative scatter correction (MSC), Fourier Transforms (FT) and Wavelet Transforms (WT) (Rinnan, Van Den Berg, and Engelsen (2009); Prakash and Y. C. Wei (2011)).

Smoothing is used to reduce noise from the hypercube without reducing the number of spectral variables. This method is usually combined with other pre-processing techniques and can be necessary as it is able to optimize the signal-to-noise ratio. Examples of the most common smoothing techniques include moving smoothing/average, median filter and Gaussian filter (Zeaiter, Roger, and Bellon-Maurel (2005)).

Derivatives, especially 1^{st} and 2^{nd} derivatives, are used to correct baseline effects in the spectra by minimising the spectral noise, reducing background effects and highlighting the differences between spectra in the hypercube. The most common derivative algorithm employed in spectral pre-processing is *Savitzky-Golay* developed by Savitzky and Golay (1964). The basic concept of the technique involves using a moving window on the data and fitting a low-degree polynomial using linear least squares. The solution to the least-squares equations can be found using analytical methods and expressed as a single set of convolution coefficients that is applied to all windows. This application then produces derivatives (estimates) at the middle point of each window which are used to smooth the spectra.

Other popular spectral pre-processing methods are the SNV and MSC, which centre and scale individual spectra, hence reducing any multiplicative interference or the spectral variability of scatter and baseline shifts. They also correct both multiplicative and additive scatter effects in the spectra of the hypercube. The FT and WT both separate noise from the spectra in the frequency domain (Burnase and Swamy (2002); Mager and Abernethy (2007)).

1.3.3 Image processing

G. M. ElMasry and Nakauchi (2016) and D. Wu and Sun (2013a) comment that many recent papers about hyperspectral imaging analysis stop once optimal wavelengths have been identified and the validated model is built, after which it is then applied to real-time applications. They argue that the general frame-

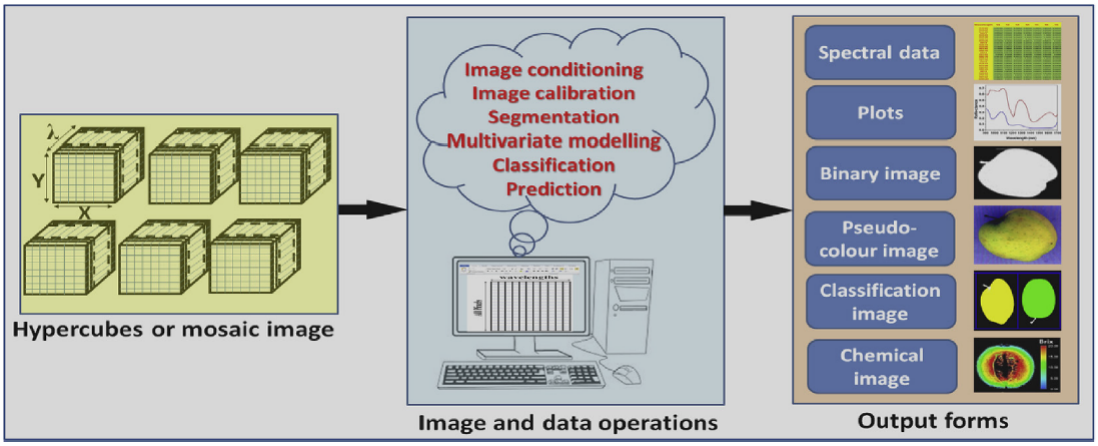


Figure 1.6: Hyperspectral analysis of an apple with image processing done to produce meaningful output. Image from G. M. ElMasry and Nakauchi (2016)

work used in conducting hyperspectral analysis should include a final step after the model has been built to reduce any misleading results due to extrapolation or interpolation when applying the model to every single pixel. This final step is image processing or post-processing. It should be used to produce an output form that can be used to enhance the visualisation for displaying important hidden information in the hyperspectral image when applied to the content distribution maps produced by the models. Various techniques can be used in the enhancement of the visualisation of the analysis such as scaling, mapping, erosion, pseudo-colour representation or dilation. Fig. 1.6 illustrates the different types of image processing operations that could be applied to the hypercubes or a collection of hypercubes to produce meaningful outputs depending on the problem that needs solving. Commonly used image processing techniques are colour mapping with intensity scaling or greyscale. This technique is used to show the difference in the component distribution between the individual pixels in the hyperspectral image. Pseudo-colour images are also widely used, where two or more images at different spectrum bands are merged to form a new image (Pohi and Van Genderen (1998)). Pseudo-colour images are used to display a greater difference between various important regions of a sample.

1.4 Outline of Thesis

As discussed previously, hyperspectral imaging has proven to be a powerful analytical tool in non-destructive food analysis. This project will explore the use of hypersectral techniques to analyse and predict the IMF content in sheep images in the visible light wavelength range. In this analysis, 101 hyperspectral sheep images were used to build and validate several multivariate mod-

els. We also explore the classification accuracy of the models when the IMF content is split into groups as defined by industry standards. Prior to prediction and classification, we investigate calculating the reflectance of each image using various methods. We then segment the images and performing spatial pre-processing to select and extract the pixels corresponding to each sheep carcass. A brief introduction of commonly used multivariate methods is given, as well as various forms that the extracted spectral data can take as input to these methods. Finally, we look at the results and discuss the outcome of the analysis.

Chapter 2

Imaging Data

In this chapter, we provide an overview of the sheep dataset and the image pre-processing steps. As described in Chapter 1, hyperspectral imaging is a technique that combines spectroscopy and imaging analysis, so as to analyse both the spectral and spatial attributes of objects. We describe the sheep images including the equipment used in acquisition of the image. We then briefly describe the processes used in calibrating the sheep images to produce reflectance values. Lastly, we pre-process the calibrated sheep images to segment between the background and the sheep carcass.

2.1 Introduction

Intra-muscular fat (IMF) content is known to play a major role in the quality attributes of meat, such as sensory properties and health considerations. The sensory attributes generally influenced by IMF content include the tenderness of meat, flavour and juiciness. A low IMF content indicates a much less appealing or tasty meat (Newcom, Baas, and Lampe (2002)).

IMF content in meat corresponds to the amount of fat within muscles , and is normally located between and within muscle fibers. It is sometimes referred to as marbling in the meat industry, and has a significant impact on marketing fresh meat. The degree of marbling is the primary determinant of the quality grade, where a higher grade means more expensive meat. Therefore, IMF content is a big indicator of meat quality and is an important factor in catergorising carcasses in the meat industry (Nunes et al. (2015)). Fig. 2.1 illustrates the relationship between the percentage of intramuscular fat content and marbling score.

Traditionally, IMF content in meat was determined using drip loss, filter paper test, cooking loss or centrifugal force. All these techniques in assessing the

Grade	Marbling score	Percentage intramuscular fat
Prime +	Abundant	
Prime °	Moderately abundant	12.3 and higher
Prime –	Slightly abundant	9.9–12.2
Choice +	Moderate	7.7–9.8
Choice °	Modest	5.8–7.6
Choice –	Small	4.0–5.7
Select +	Slight +	3.1–3.9
Select –	Slight –	2.3–3.0
Standard +	Traces	2.2 and lower
Standard °	Practically devoid	
Standard –	Practically devoid –	

Figure 2.1: Relationship between IMF content, Marbling score and carcass quality grade. Adapted from Lonergan, Topel, and Marple (2018)

IMF content in meat are time consuming, destructive and in most cases unsuitable for industry applications (Kamruzzaman, G. ElMasry, et al. (2012a). This prompted the development of modern techniques to estimate IMF content in meat. Nowadays, IMF can be measured by various analytical methods such as spectroscopy (Hocquette et al. (2010)) or in real time via the use of ultrasound (Hassen et al. (2001), Newcom, Baas, and Lampe (2002)) and H^1-NMR techniques (Boesch et al. (1997)). Recently, spectroscopy and computer imaging have been used to estimate the IMF content in meat. Individually the techniques are not very accurate, since spectroscopy techniques such as near-infrared spectrometry only measure the spectral properties of the meat, while computer imaging measures the spatial properties. One technique that has shown promise in estimating the IMF content is hyperspectral imaging because it combines the use of spectroscopy and computer imaging analysis. It has been shown to have good precision in recent work on pork and beef (Kamruzzaman, G. ElMasry, et al. (2012b)). The use of hyperspectral imaging has been favourable to traditional techniques given its non-invasive, non-destructive and efficient process.

In this thesis, we use three software packages in the reading, calculation and analysis of the hyperspectral images. Sycven, a software package devel-

Figure 2.2: Hyperspectral imaging acquisition stage. Left panel shows the setup of the Hyperspectral equipment. Right panel shows the recovered pseudo-colour image of a sheep carcass. Adapted from Robles-Kelly and R. Wei (2015)

oped by Scyllarus CSIRO for analysing hyperspectral images, was used in the reflectance calculation of the sheep images. Matlab was used to convert the images to text files using a toolbox develop by Scyllarus CSIRO (Habili and Oorloff (2015)). All statistical analyses and image segmentation were conducted in R, where models were trained and tested using the *caret* package (Kuhn (2008)).

2.2 Data

One hundred and one hyperspectral images of sheep carcasses were acquired at the WAMMCO Katanning abattoir in Katanning, Western Australia and was provided by Dr Graham Gardner and Dr Khama Kelman. The sheep carcasses were cut in such a way to show a cross-section of the eye muscle. Using two tungsten lamps for illumination, the sheep carcass was positioned in front of the camera. The left panel of Fig 2.2 shows the set up used for acquiring the hyperspectral image, while the right panel in the same figure, shows the recovered pseudo-colour image of a sheep. That is, the hyperspectral image was rendered in colour.

The camera used in the acquiring of the sheep images, as depicted in Fig 2.2, was an OKSI Turnkey system based upon liquid crystal tunable filters (LCTFs) with a spatial resolution of 2Mp. The camera produces hyperspectral images in the visible wavelength range, that is, a wavelength range between 400nm and 700nm, with steps of 10nm (Robles-Kelly and R. Wei (2015)). This results in a hyperspectral image with 31 different wavelengths per sheep, giving in total 3131 images in our dataset. Each image comprised of 1,447,680 pixels (1392 x

1040).

The rest of this chapter describes in detail the reflectance calculation and spatial pre-processing steps used on this dataset. Section 2.3 discusses the methods used to calculate the reflectance. Section 2.4 illustrates our choice of illuminants in the reflectance calibration. Finally, section 2.5 discusses the methods used in spatial pre-processing of the images.

2.3 Reflectance Calibration

As discussed in section 1.3.1, calculating the reflectance values of an image decreases the effect of the residual electric current from the camera and minimizes the influence of the background spectral response. The process to obtain the reflectance values from the acquired raw image involves the use of two reference images, that we will define as *dark* and *white* image. The *dark* image is a reference image with approximately 0% reflectance that is acquired by placing the camera's opaque cap on its lens. Similarly, the *white* image is a reference image with approximately 100% reflectance that is acquired by placing a uniform white ceramic tile over the camera's lens. It is also sometimes referred to as the background image (Kamruzzaman, Makino, and Oshita (2016)). With the two reference images acquired, the reflectance values (R) of the raw hyperspectral image are calculated using the following formula: (G. ElMasry, Sun, and Allen (2013)):

$$R = \frac{Object - Dark}{White - Dark}.$$

This process is popular in real world applications given the ease of acquiring the reference images. However, in some cases acquiring the reference images may not be possible due to unforeseeable circumstances as is our case in this thesis. In such cases there are other methods (Rahman and Robles-Kelly (2013)). In the next section, we look at the dichromatic model, a method developed by Shafer (1985) and adapted by a group at National ICT Australia (now part of CSIRO), which we refer to as NICTA throughout.

2.3.1 Dichromatic Model

The dichromatic model assumes that the total radiance L_t of the reflected light is the sum of two independent parts: the radiance L_j of the light reflected at

the interface and the radiance L_b of the light reflected from the surface body (Shafer (1985)). That is:

$$L_t(\lambda, j, e, h) = L_j(\lambda, j, e, h) + L_b(\lambda, j, e, h), \tag{2.1}$$

where λ is the wavelength of light, j the angle of incidence between the illumination direction and the surface normal (i.e. direction perpendicular to the surface of the object), e the angle of emittance representing the angle between the surface normal and the viewing direction and h the phase angle representing the angle between the illumination and viewing directions.

The NICTA adaptation of this dichromatic model states that the light reflected at the interface can be denoted by surface reflectance and the radiance of the light reflected from the surface of the body denoted by surface radiance. The adaptation involves decomposing the surface radiance into a diffuse and specular component (Huynh and Robles-Kelly (2010)), resulting in the dichromatic model

$$I(\lambda,\ i) = g(i)\, L(\lambda)\, S(\lambda, i) + k(i)\, L(\lambda). \tag{2.2}$$

This is achieved by letting an object with surface radiance $I(\lambda,\ i)$ at pixel i and wavelength λ be illuminated by an unknown illuminant whose spectrum is $L(\lambda)$. The other parameters of the model include the shading factor $g(i)$ which governs the percentage of diffuse light and depends solely on the surface geometry, specular coefficient $k(i)$, light spectrum $L(\lambda)$ and the spectral reflectance $S(\lambda,\ i)$ at pixel i and wavelength λ from the spectral radiance $I(\lambda,\ i)$.

To recover the surface reflectance for further analysis, (2.2) can be rearranged to form:

$$S(\lambda,\ i) = \frac{1}{g(i)} \left(\frac{I(\lambda,\ i)}{L(\lambda)} - k(i) \right). \tag{2.3}$$

It can be noted that, if the specular coefficient $k(i)$, shading factor $g(i)$ and illuminant power spectrum $L(\lambda)$ are readily available, it is easy to compute the surface reflectance of the object using (2.3). However, in some instances these components are not readily available and would need to be recovered by applying methods such as those presented in Huynh and Robles-Kelly (2010). A more in-depth description of parameter recovery, as well as the optimal approach is detailed in Huynh and Robles-Kelly (2008).

In order to calculate the surface reflectance we need to first estimate the illuminant power spectrum $L(\lambda)$. Therefore, in the next section 2.3.2 we look at

various ways to estimate the illuminant power spectrum. Section 2.3.3 gives a brief overview of the approach taken by NICTA in the estimation of $L(\lambda)$ as well as the other parameters in the (2.3).

2.3.2 Illuminant Estimation

The illuminant power spectrum $L(\lambda)$ is a mathematical representation of a theoretical light source. It is used in the calculation of tri-stimulus values from a spectrophotometric measurement. That is, the calculation of the light intensity based on the three primary colour values (Red, Green, Blue). The tri-stimulus values represent the relative power of the theoretical source for specific wavelengths in the visible spectrum. The use of the theoretical representation of the light source is mainly due to the relative power distribution of a real light source, which may be difficult to standardize.

For the initialisation of the illuminant power spectrum $L(\lambda)$ we briefly consider the following algorithms: Finlayson and Schaefer (Finlayson and Schaefer (2001)), grey world (Buchsbaum (1980)), white patch (W.-R. Lee, Hwang, and Jun (2011)), shades of grey (Finlayson and Trezzi (2004)) and grey edge (Van De Weijer, Gevers, and Gijsenij (2007)).

Finlayson and Schaefer

The method proposed relies on the dichromatic plane hypothesis. That is, the model implies a two-dimensional colour space of pixels in patches with homogeneous reflectance. Illumination estimation is viewed as an optimisation problem to maximise the total projection length of the light colour on all dichromatic planes. As such, the approach predicts the illuminant as the intersection of the dichromatic planes. The approach was developed on trichromatic images, but can be adapted to multispectral/hyperspectral images by providing homogeneous surface patches to the colour constancy algorithm and then solving the eigen-system of the sum of the projection matrices on the dichromatic planes. The resulting eigenvector with the largest corresponding eigenvalue is deemed to be the illuminant for that patch (Finlayson and Schaefer (2001)).

Grey World

Grey world is a renowned colour-constancy method that assumes, on average, the world is grey (Buchsbaum (1980)). This means that the average value of each wavelength in an image will be close to $\frac{1}{2}$, when calculated using the

following formula:

$$\lambda_i = \frac{1}{n}\sum_{x,y} J(x,y) = L_i \frac{1}{n}\sum_{x,y} G\,(x,y)\, R_i(x,y), \tag{2.4}$$

where i is the number of wavelengths in the image, n is the number of pixels, λ_i is the i^{th} wavalength and J is the input image, that is acquired by an imaging device with the corresponding formula:

$$J\,(x,y) = G\,(x,y)\, R_i\,(x,y)\, L_i\,(x,y)\,, \tag{2.5}$$

with $G\,(x,y)$ a factor due to the scene geometry at position (x,y), $R_i\,(x,y)$ the reflectance at the position (x,y) and $L_i\,(x,y)$ the illuminant at position (x,y).

Taking the expectation of GR, when considering the range of the reflectance to be [0, 1],

$$E\,[GR_i] = E\,[G]\,E\,[R_i] = E\,[G]\left(\int_0^1 x\,dx\right) = E\,[G]\,\frac{1}{2} \tag{2.6}$$

Therefore, taking the expected value of each wavelength λ_i, we get:

$$\lambda_i = \; L_i \frac{1}{n}\sum_{x,y} G\,(x,y)\, R_i(x,y) \approx L_i E\,[GR_i] = \; E\,[G]\,\frac{1}{2} \tag{2.7}$$

And the illuminant L_i as well as the output image $o_i(x,y)$ could be estimated using:

$$L_i \approx \frac{2}{E\,[G]}\lambda_i o_i\,(x,y) = \frac{J_i\,(x,y)}{L_i} \approx G\,(x,y)\, R_i\,(x,y)\,. \tag{2.8}$$

White Patch

The white patch method is a simplified version of the Retinex method with the aim to compute the luminosity of an image using logarithms and iterations (W.-R. Lee, Hwang, and Jun (2011)). The difference between the two methods is that white patch assumes that the scene includes a white patch and that the light source affects the whole scene equally. As such, the illuminant is estimated by selecting the maximum value per wavelength.

$$L_{i,max} = \max_{x,y}\{J_i(x,y)\}, \tag{2.9}$$

since the patch would reflect the maximum of the incident light at each wavelength. The output image can be reproduced using (2.8) with $L_{i,max}$ instead of L_i.

Shades of Grey

The shade of grey method follows from both the grey world and white patch method being forms of the Minkowski-norm:

$$\left(\frac{\int (f(x))^p dx}{\int dx} \right)^{\frac{1}{p}} = ke, \tag{2.10}$$

where k is a multiplicative constant, $p \in [0, \infty]$ and e is the vector of the light source colour. Finlayson and Trezzi found that the shades of grey method can be represented by the Minkowski-norm, with a p value of 6. Coincidentally, they found that when p is 1, (2.9) had similar results to the grey world method, whereas, when p is ∞, the results are similar to the white patch method (Finlayson and Trezzi (2004)). The shades of grey method shares the same properties of both the white patch and grey world methods.

Grey Edge

Another renowned state-of-the-art method in colour constancy is the grey edge method. This methods works on the assumption that the average of the reflectance differences in a scene is achromatic. It can also be represented as a Minkowski-norm:

$$\left(\int \left| \frac{\partial^n f^\sigma(x)}{\partial x^n} \right|^p dx \right)^{\frac{1}{p}} = ke^{n,\, p,\sigma}, \tag{2.11}$$

where the parameters, $\{n,\ p,\ \sigma\}$, for the grey edge are $\{1, 1, 6\}$.

2.3.3 NICTA Parameter Estimation

Using the estimation of the illuminant power spectrum $L(\lambda)$ from any of the methods described in section 2.3.2 and (2.2), the parameters $g(i)$, $k(i)$ and $S(\lambda, i)$ can be estimated.

The parameter estimation can be achieved by minimising the dichromatic cost function $F(I)$ of a hyperspectral image I, where the cost function is the weighted sum of the dichromatic error and a regularisation term $R(i)$ for each image location (Huynh and Robles-Kelly (2008)). That is,

$$F(I) = \sum_{i \in I} \left[\sum_{j=1}^{n} \left[I(\lambda_j,\ i) - L(\lambda_j) \left(g(i)\, S(\lambda_j,\ i) + k(i) \right) \right]^2 + \alpha R(i) \right], \tag{2.12}$$

where α is a constant that acts as a balancing factor between the regularisation term $R(i)$ and the dichromatic error.

Before defining the regularisation term $R(i)$, Huynh and Robles-Kelly (2008) notes that without any further constraints, minimising the cost function pro-

duces an underdetermined problem. Given that an image that has n wavelengths with m pixels each, we would need to minimise over $2m + n + (m \times n)$ variables, however there are only $m \times n$ terms in the cost function $F(I)$. By further constraining the model to smooth surfaces made of the same material, two constraints can be imposed on the cost function.

Firstly, a common diffuse reflectance is shared across all locations on the surface. Therefore, it is assumed that a uniform albedo surface, that is the proportion of light reflected by a surface, P will have the same reflectance for each pixel, i.e. $i \in P,\ S(\lambda_j, i) = S_P(\lambda_i)$. This constraint reduces the number of unknowns $S(\lambda_j, i)$ from $m \times n$ to $N \times n$, where N denotes the number of surface albedos in the scene.

The second constraint involves applying a smoothing variation on the geometry of P that allows the derivation of the regularisation term $R(i)$ as a function of the shading factor $g(i)$ and is not dependent on the wavelength. This implies that the shading factor $g(i)$ should vary smoothly across P and can be formulated by minimising the variation of the gradient magnitude of the shading map. Effectively, preventing discontinuities in the shading map of P,

$$R(i) = \left[\frac{\partial g(i)}{\partial x(i)}\right]^2 + \left[\frac{\partial g(i)}{\partial y(i)}\right]^2 \tag{2.13}$$

where $x(i)$ and $y(i)$ are the coordinates for the pixel at location i. Thus, by imposing the first and second constraints mentioned previously on the cost function the dichromatic model parameters can be recovered by minimising $F^*(I)$ as an alternative to $F(I)$, where:

$$F^*(I) = \sum_{p \in P} F(P) \tag{2.14}$$

$$= \sum_{p \in P} \sum_{i \in P} \left[\sum_{j=1}^{n} \left[I(\lambda_j,\ i) - L(\lambda_j) \left(g(i) S_P(\lambda_j) + k(i) \right) \right]^2 + \alpha R(i) \right]. \tag{2.15}$$

To find parameters $L(\lambda_j)$, $S_P(\lambda_j)$, $g(i)$ and $k(i)$ that would yield the minimum of the cost function, an iterative approach is adopted. Defining:

- $\mathbf{L} = [L(\lambda_1), \ldots, L(\lambda_n)]^T$ spectral power vector of the illuminant,
- $\mathbf{S}_P = [S_P(\lambda_1), \ldots, S_P(\lambda_n)]^T$ common spectral reflectance vector for each surface P,
- $\mathbf{g}_P = [g(i_1), \ldots, g(i_l)]^T$ shading map of all pixels in surface P, with $i_1, \ldots i_l$ being all the pixels in the surface P

- $\mathbf{k}_P = [k(i_1), \ldots, k(i_l)]^T$ specularity map of all pixels in the surface P

It can be seen that at each iteration t, the cost function can be minimised with respect to $\mathbf{L}$ and $\{\mathbf{g}_P, \mathbf{k}_P, \mathbf{S}_P\}$ in separate steps for each P.

The dichromatic variables $\{\mathbf{g}_P, \mathbf{k}_P, \mathbf{S}_P\}$ can then be optimised with respect to the image radiance and the illuminant $\mathbf{L}$, when indexed to t iterations. This suggests that once the variables $\{\mathbf{g}_P, \mathbf{k}_P, \mathbf{S}_P\}$ are at hand, then it is possible to obtain a closed form solution for $\mathbf{L}$.

The method is then decomposed into two interleaved minimisation steps. It begins with the initialisation of the light spectrum $L(\lambda)$. This can either be the illuminant power spectrum $L(\lambda)$ estimated by the methods in section 2.3.2 or by the using Huynh and Robles-Kelly (2008) initialisation, which employs the light spectrum estimated by the brightest pixel per band. They argue that the radiance of a surface cannot be greater than the irradiance of the illuminant, therefore the highest radiance across the pixels for each wavelength can be used as a lower bound estimate of the light spectrum.

The first step of the interleaved minimisation steps involves estimating an optimal surface reflectance and geometry using the light spectrum $\mathbf{L}^{t-1}$. This step, therefore, is reduced to minimising

$$F(P)|_{\mathbf{L}^{t-1}} = \sum_{i \in P} \left[\left\| I(i) - g(i) D_P^{t-1} - k(i) \mathbf{L}^{t-1} \right\|^2 + \alpha R(i) \right], \tag{2.16}$$

where $\mathbf{I}(i) = [I(\lambda_1, i), \ldots, I(\lambda_n, i)]^T$ is the spectral radiance vector at image pixel i, $\mathbf{D_P^{t-1}} = < \mathbf{L}^{t-1}, \mathbf{S}_P >$ is the L_2-norm of the illuminant power vector and the spectral reflectance vector that defines the diffuse radiance components for the surface P and $\|.\|$ denotes the inner product. This shows that at iteration t, the unknowns $\mathbf{g}_P^t$, $\mathbf{k}_P^t$ and $\mathbf{S}_P^t$ can be solved separately for each surface P.

In the second step of each iteration t, $\mathbf{L}^t$ can be solved, given $\mathbf{g}_P^t$, $\mathbf{k}_P^t$ *and* $\mathbf{S}_P^t$, in a closed form as

$$\mathbf{L}^t(\lambda) = \frac{\sum_{p \in P} \sum_{i \in P} I(\lambda, i) \left(g^t(i) S_p^t(\lambda) + k^t(i) \right)}{\sum_{p \in P} \sum_{i \in P} \left(g^t(i) S_p^t(\lambda) + k^t(i) \right)^2}. \tag{2.17}$$

It is noted that the second term $R(i)$ in the cost function disappears, this is due to it being independent of the illuminant. The closed form solution can also be found by differentiating equation (2.15) with respect to the illuminant irradiance and equating the resulting expression to zero.

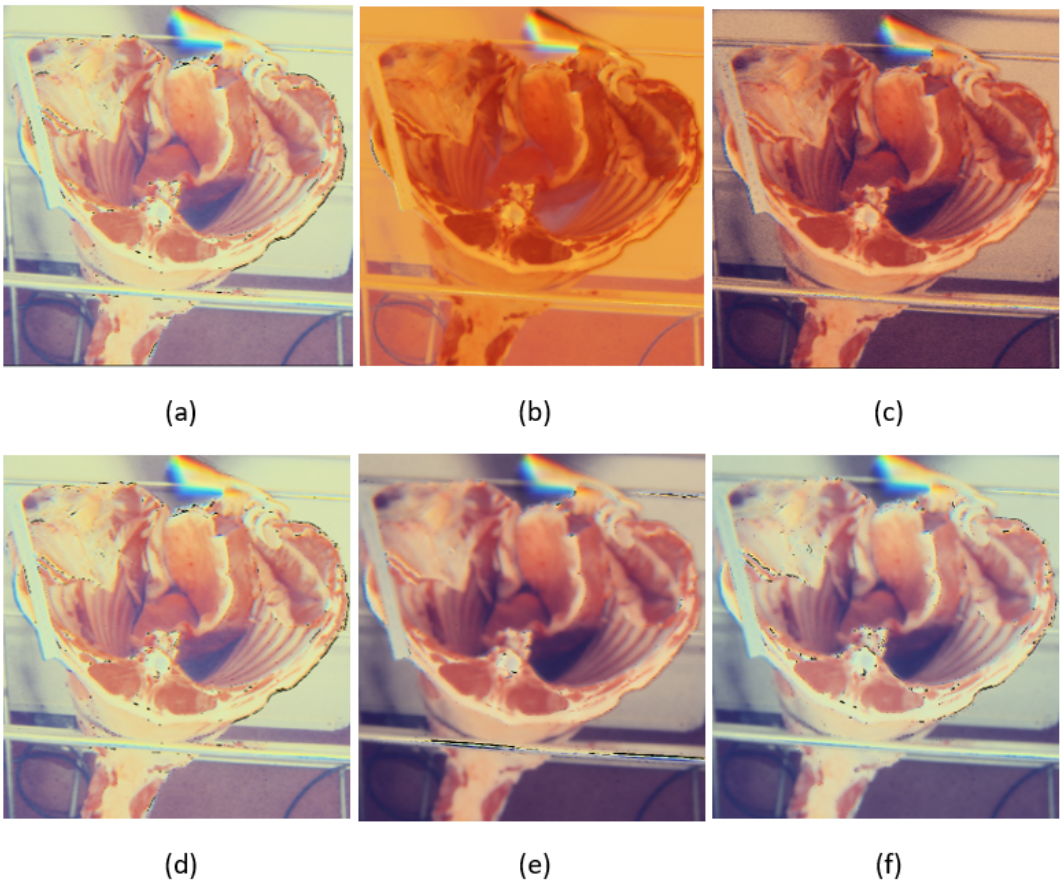


Figure 2.3: Recovered pseudo-colour image of a Sheep carcass using (a) Grey World (b) White Patch (c) NICTA (d) Shades of Grey (e) Grey edge and (f) Finlayson Shaefer. *Images are cropped*

2.4 Illuminant Choice

There are no known advantages or disadvantages in using one illuminant recovery method over another. However, we explored the different methods introduced in section 2.3.2, to determine which might be most suitable for the dataset under consideration.

Scvyen was used to recover the reflectance using each of the six illuminant recovery methods mentioned in sections 2.3.2 and 2.3.3. That is, NICTA, Finlayson and Schaefer, Grey edge, Grey World, Shades of Grey and white patch. Fig 2.3 shows the recovered pseudo-images for each of the different illuminant recovery methods for a sheep carcass chosen at random from the dataset.

To determine if there is a difference between the illuminant recovery methods, we calculated the correlation between each of the methods. Fig 2.4 shows a plot of the pairwise Pearson correlations between the various methods. Overall, from the figure we note that all possible combinations of methods have a correlation of 85% or greater for each wavelength. This suggests that the meth-

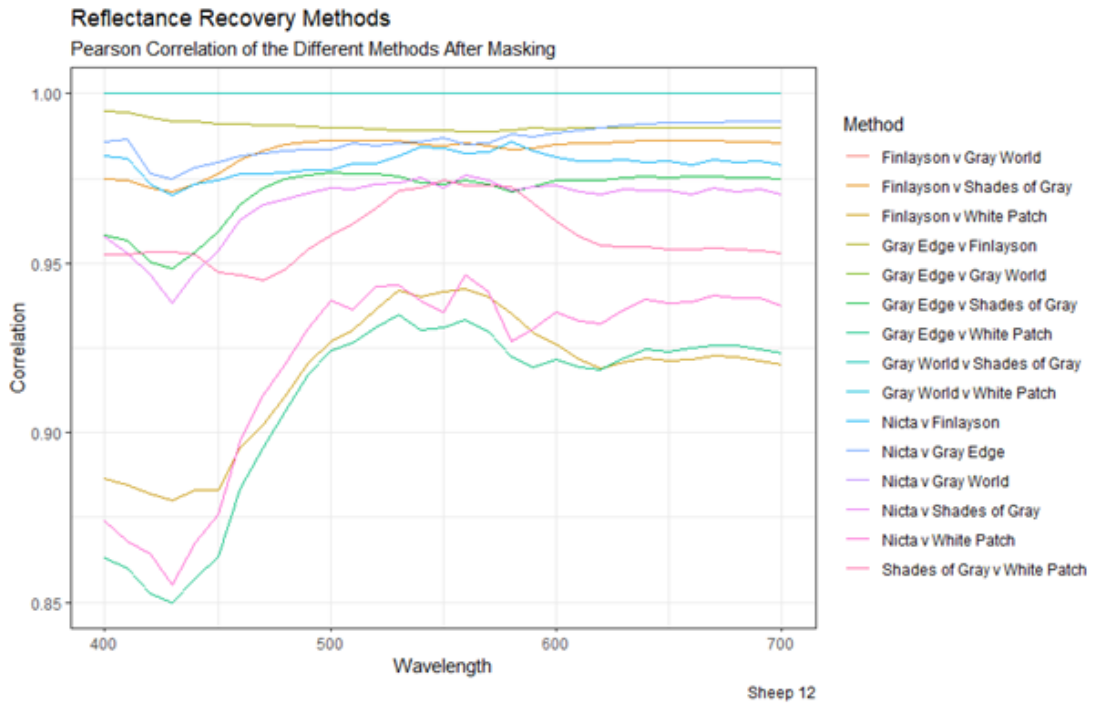


Figure 2.4: Correlation graph for all possible combinations of the 6 different methods Grey World, White Patch, NICTA Reflectance, Shades of Grey, Grey Edge and Finlayson and Shaefer

ods are highly correlated in this sheep dataset. Alternatively, using the Spearman correlation (see Appendix Fig 5.1), it can again be seen that the methods are highly correlated, with correleation coefficient of at least 96%. We also note that there appears to be no difference between the grey edge and shades of grey methods (correlation = 1) across each wavelength. In contrast, the grey edge and finlayson and schaefer method are strongly correlated with a correlation of more than 97%. On the basis of these results, the default method in Scyven, NICTA, is adopted for the illuminant initialisation, as well as the reflectance calculation for the rest of the analysis in this thesis and for the prediction of the intramuscular fat content and classification of the marbling score in the sheep carcasses.

2.5 Spatial Pre-processing

Spatial pre-processing refers to pre-processing done to an image prior to conducting any analysis. In our case, it involves the identification of the sheep carcass, the region of interest (ROI), from the image. That is, creating a matrix with the same pixel dimension as the sheep image and setting the location of the pixels denoting the sheep carcass to *1* and the rest to *0*. Once the ROI has been identified, the original sheep image as well as the hypercube can be masked, leaving only the ROI and removing the background. This is usually

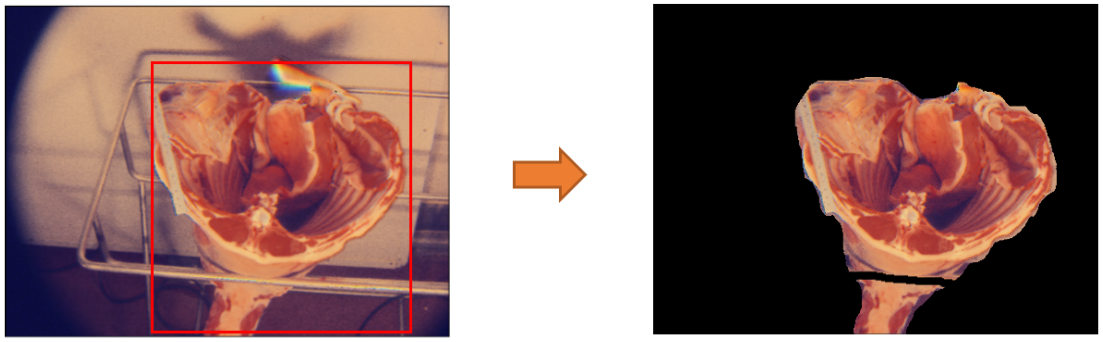


Figure 2.5: Illustrates the masking process. Left the pseudo-colour image and the ROI. Right same image with only the ROI selected and the rest as background

a necessity in hyperspectral imaging since there may be various objects in the image and the researcher is often be interested in only some of them.

In Fig. 2.5 we illustrate the masking process of identifying and selecting our ROI. In the left panel, we show the pseudo-colour image and the area in the red square which delineates our ROI. In the right panel, we have extracted the spectral features of the image in the ROI and rendered a pseudo-colour image to display the ROI we are interested in.

There are various methods that can be employed in the selection of the ROI. We, however, found the use of k-means, an unsupervised clustering method, to be consistent and worked very well in identifying the ROI without further editing of the masked images. We compared using 2, 4, 6 or 8 clusters to determine which would be able to successfully segment the sheep from the background in the image. In Fig. 2.6, we show a visual representation of the different clusters.

We note, from Fig 2.6, that the clustering algorithm K-means with 2 clusters (B) is not able to accurately distinguish between the background and the ROI and would therefore not provide a good mask for the hypercube. (C) and (D), 4 and 6 clusters respectively, provide a fairly good mask that can be applied on the hypercube, with (D) showing a better mask. While 8 clusters (E) provides a more detailed image than the rest of the values, the drawback is the computational cost. At the same time, K-means with 4-6 clusters produces a fairly good mask with lower computational cost. Therefore, we consider between 4 and 6 clusters when masking the rest of the images for the selection of the ROI in this thesis.

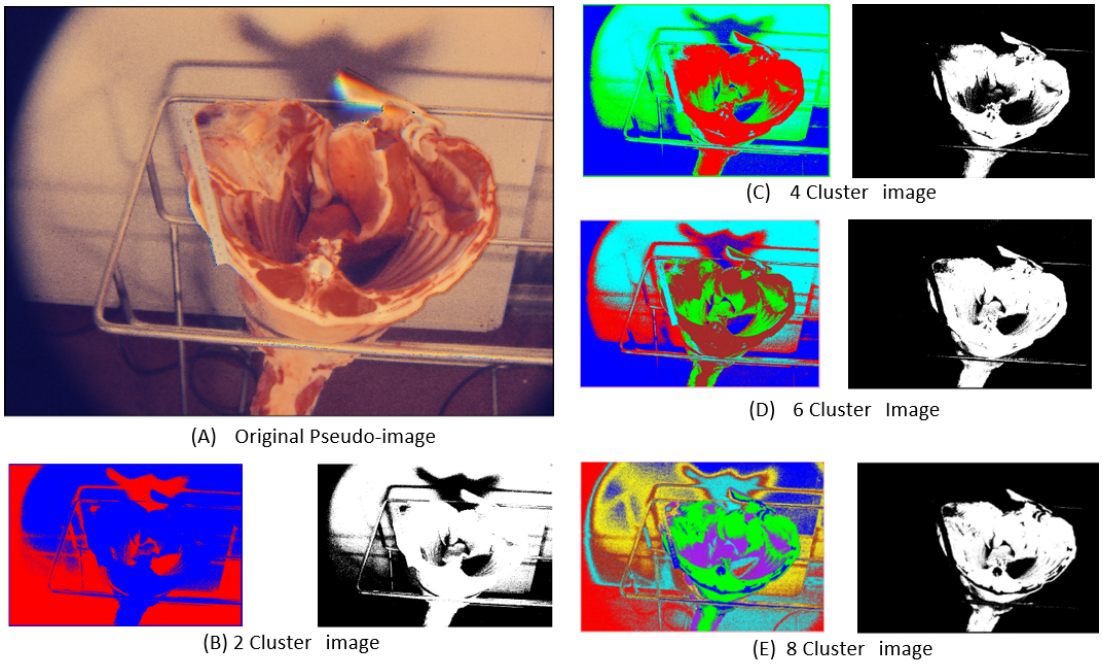


Figure 2.6: Showing the original Pseudo-image (A) as well as Cluster image and corresponding binary image of 2 cluster (B), 4 cluster (C), 6 cluster (D) and 8 cluster (E)

Chapter 3

Multivariate Analysis

In this chapter, we provide an overview of the IMF data corresponding to the sheep images. We propose three formats in which data can be extracted from the segmented sheep images. We also briefly provide an overview of some of the common multivariate methods used in hyperspectral imaging that we will employ in the analysis of the sheep images and IMF data. These methods include multiple linear regression, principal component analysis and regression, partial least squares regression and discriminant analysis, linear discriminant analysis, support vector machines and artificial neural networks. Lastly, we provide an overview of how we can measure model performance.

3.1 Introduction

The hyperspectral image contains a wealth of data that are commonly extracted as texture-based, morphological-based (Shape) and intensity-based features. The use of multivariate techniques is usually required to decompose the features contained in the hyperspectral image into valuable information. This information can then be used to explore the desired object's characteristics/attributes. The use of multivariate as opposed to univariate techniques is preferred as multivariate techniques are capable of simultaneously analyzing data collected on more than one variable, especially when two or more variables are not orthogonal to each other, i.e. there is high correlation between variables. Some multivariate techniques are also capable of variable selection or dimension reduction. That is, they either select a small number of wavelengths that are able to explain the hypercube (i.e. remove any redundant wavelengths) or group wavelengths that share the same characteristics (i.e. reduce the number of dimensions considered during the model creation).

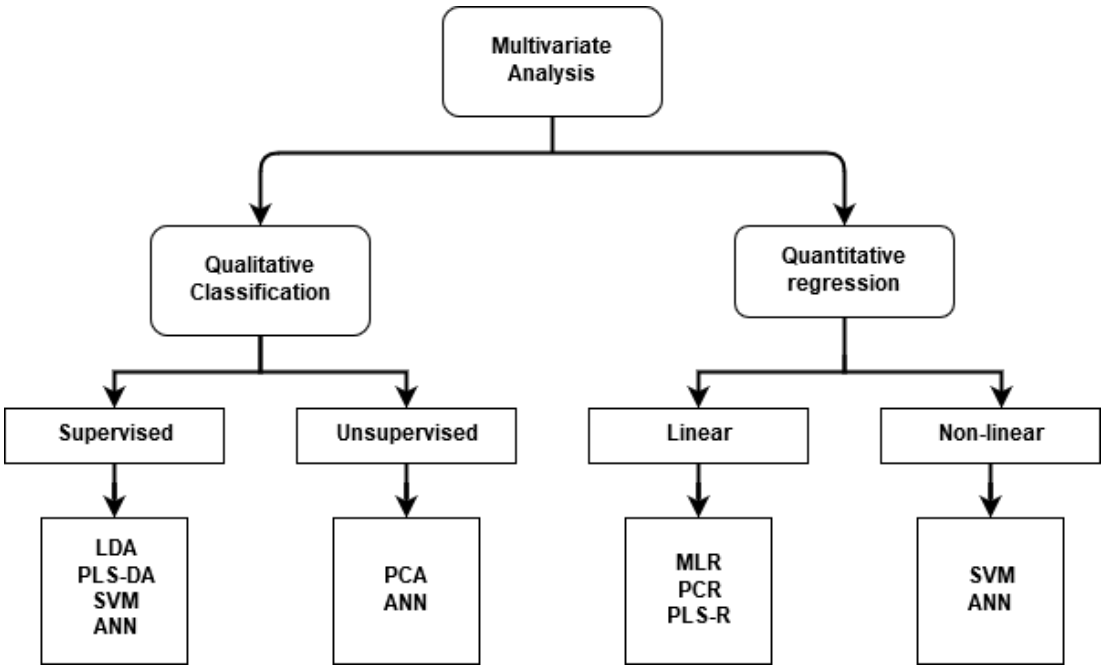


Figure 3.1: Classification of multivariate analysis techniques into methods widely employed in hyperspectral data analysis

Generally, multivariate methods can be classified into quantitative regression or qualitative classification as illustrated in Fig. 3.1. Prior to the use of any quantitative or qualitative techniques, the hypercube needs to be in a format that can easily be used. Fig. 3.2 illustrates the change from the hypercube to a spectral data matrix with each row representing a different wavelength.

The general aim of using quantitative regression methods is to build a model that is able to establish a relationship between the object's numeric and continous attributes to the wavelength and/or spatial information contained in

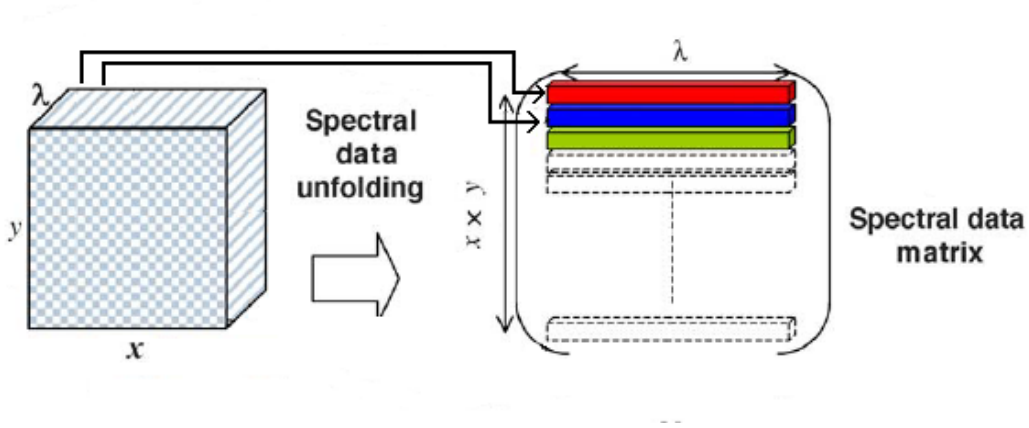


Figure 3.2: Spectral data unfolding of a hyperspectral hypercube to a spectral data matrix that can be used as an input into the multivariate techniques. Adapted from Roggo et al. (2005)

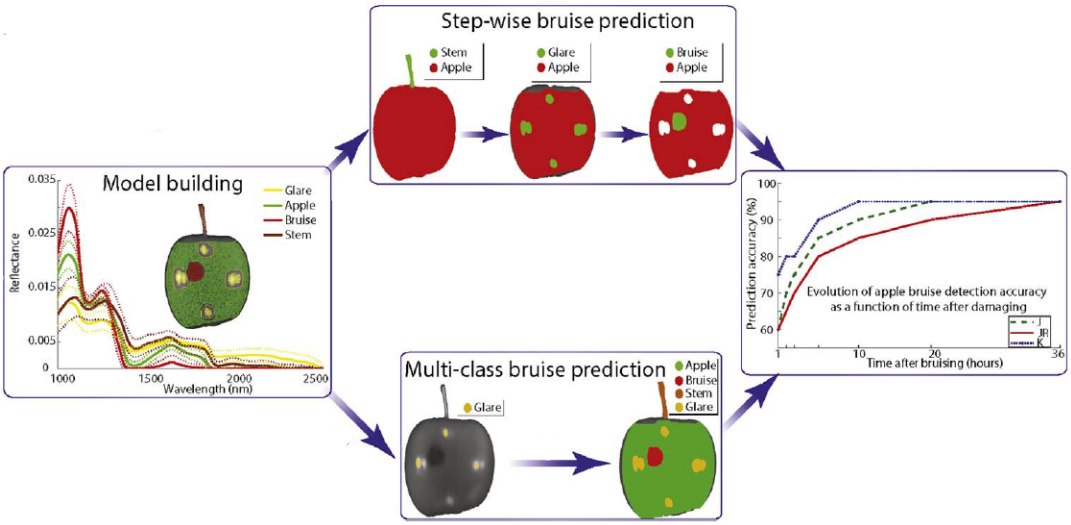


Figure 3.3: Hyperspectral image analysis on an apple. Adapted from Keresztes et al. (2017)

the hypercube. Furthermore, quantitative methods can be either linear or non-linear, with multiple linear regression (MLR), principal component regression (PCR) and partial least squares regression (PLS-R) being the most widely used linear methods and support vector machines (SVM) with non-linear kernels or artificial neural network (ANN) used for non-linear models.

Alternatively, the purpose of qualitative classification is to build a model using the wavelength and/or spatial information contained in the hypercube to classify certain binomial or categorical characteristics of the sample. Qualitative classifications can be further classified as supervised and unsupervised. Supervised classification maps an input (spectra and/or spatial information contained in the hypercube) to a desired output (target characteristic) based on given sample input-output pairs. Methods include linear discriminant analysis (LDA), SVM, ANN and PLS with a discriminant analysis extension (PLS-DA). Unsupervised classification uses only the input (wavelength and/or spatial information contained in the hypercube) with no corresponding output defined, hence the classifier generally tries to find patterns using the given information, methods include principal component analysis (PCA) and ANN clustering.

To illustrate the use of quantitative and qualitative techniques on hyperspectral analysis, consider an apple used in a study conducted by Keresztes et al. (2017), displayed in Fig. 3.3. The qualitative approach of model creation would result in a model able to classify between different objects in the image. The model would also be able to determine whether or not there was a bruise (defect) in a given apple, as illustrated by the step-wise or multi-class bruise

prediction on the figure. The quantitative approach would result in a model that is able to predict the time after bruising for the apple, the firmness of the apple, the moisture content within the apple and/or the chemical attributes of the apple.

In this thesis, our overall aim was to predict the IMF content of sheep using hyperspectral image analysis. To do this, we processed the sheep images as depicted in Chapter 2. Now we would like to see if the processed sheep image can be used to predict the IMF content using some of the multivariate methods mentioned above.

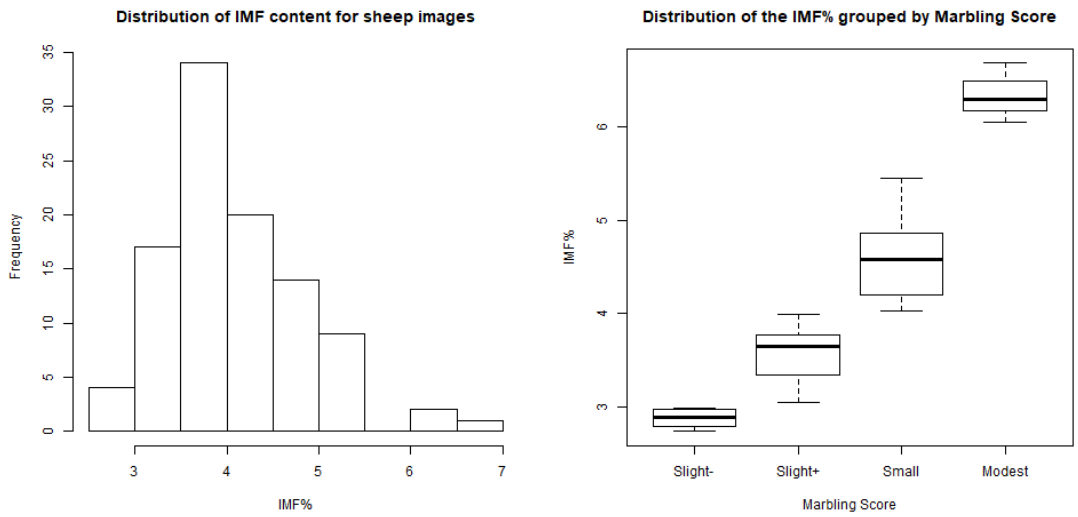
In the following sections, we let $\mathbf{Y}$ denote a $(n \times q)$ matrix of response (dependent/outcome) variables, the object's attributes of interest is the hyperspectral image, in our case the IMF content (quantitative) or the marbling score (qualitative). Let $\mathbf{X}$ denote a $(n \times p)$ design matrix of explanatory (independent) variables, i.e. the spectral bands of the sheep hypercube. Let β denote a $(p \times 1)$ coefficient matrix corresponding to the explanatory variables and ϵ denote a $(n \times q)$ matrix of errors that are independent identically distributed normal with a mean of 0 and a variance-covariance matrix of $\sigma^2\mathbf{I}_n$, where $\mathbf{I}_n$ is the identity matrix. Note that q is the number of the object's attributes to be considered, p is the number of spectral values in the hypercube and n is the total number of observations.

3.2 Data Matrix

Before discussing the various methods that can be applied to the data in order to predict the intra-muscular fat content or classify the marbling score, we first consider the format the data needs to be in. We investigated several alternative ways to extract information from the sheep hypercube into the design matrix $\mathbf{X}$, described below in more details. We also provide summary statistics of the IMF content data in our response variable $\mathbf{Y}$.

3.2.1 Intramuscular Fat content

In order to perform any analysis on the sheep images, corresponding IMF content values are required. Fig. 3.4 (a) shows the distribution of the 101 IMF content values of the sheep in our dataset. The distribution of the IMF content values appears slightly skewed with a right-tail. From this, we see that the vast majority of sheep have IMF content values between 2.5 and 5.5, with very few sheep with IMF content values in the range of 6 to 7. The IMF content values seem to be centered around 4%. Table 3.1 shows the exact summary statistics



(a) **(b)**

Figure 3.4: Histogram of the IMF content shows a slightly right-tailed distribution.

in the total row for the IMF content values in the dataset.

We can group the IMF content values into 4 categories corresponding to marbling scores of Slight–, Slight+, Small and Modest. Fig 3.4 (b) shows the distribution of the IMF content values grouped by the marbling score. For the Slight– group, the figure shows a symmetrically distributed boxplot with mean and median of approximately 2.8 as depicted by Table 3.1. Similarly, for the Modest group with mean and median of approximately 6.3. The Slight+ and Small group appear to have slighlty different means and medians, with the former having a mean of approximately 3.5 and a median of approximately 3.6 and the latter having a mean of 4.6 and median of 4.5.

Table 3.1: Descriptive statistics for the IMF content in the Dataset

		Intramuscular Fat Content				
		n	**Mean**	**Median**	**SD**	**Range**
	Slight–	4	2.847	2.885	0.114	0.525
Marbling	Slight+	51	3.587	3.652	0.242	0.937
Score	Small	43	4.612	4.582	0.438	1.433
	Modest	3	6.349	6.304	0.317	0.629
	Total	101	4.077	3.882	0.756	3.950

The Modest and the Slight– have less than 5 IMF content observations. Unfortunately, classification models would not be capable of accurately classifying these two groups due to the small samples. We therefore decided to merge the Slight– group with the Slight+ group and the Modest group to the Small group. We were therefore able to explore predicting the actual IMF content values (quantitative variable, continous value) or the marbling score (*0/1*, that is Small vs Modest, qualitative variable) in this dataset.

3.2.2 Column-Average

The first format we propose involves taking the average of the ROI for each wavelength after spatial pre-processing of the hypercube. That is we simply take the column average in the spectral data matrix (see Fig. 3.2). The resulting data matrix would be $1 \times n$, where n denotes the number of wavelengths in the hypercube. When combining all images in the dataset, the resulting matrix would be 100×31, i.e. 100 images and 31 wavelengths.

This data matrix is the easiest to compute of the alternatives considered here.

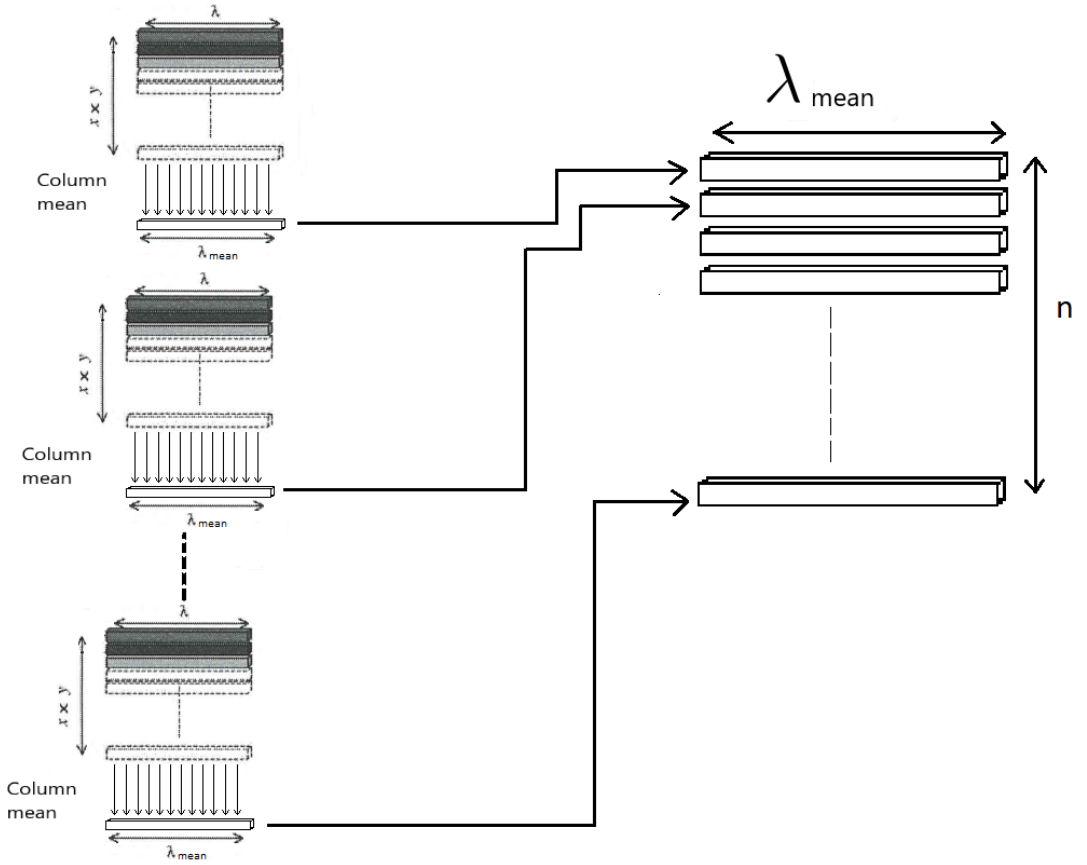


Figure 3.5: Visualisation of the Column-Average data matrix. Formed by taking the average for each column in the spectral data matrix . Adapted from Roggo et al. (2005)

It is also the simplest, but at the cost of losing information due to taking the mean. An advantage over the other two formats would be, the identification of optimal wavelengths. Fig. 3.5 visualizes the data extraction process. This data format is also widely used in real world applications when building models for the prediction or classification of an attribute, such as the IMF content or marbling score.

3.2.3 Row-Average

The second format we propose involves taking the average across the wavelengths for every pixel after selecting and extracting information from the ROI in the hypercube. That is, we simply take the row average in the spectral data matrix (see Fig. 3.2). The resulting data matrix would be $m \times 1$, where m is the number of pixels in the ROI. When combining all images into one dataset, the resulting matrix would ideally be $m \times 101$. However due to different sheep images having different numbers of pixels in the ROI, we instead sample a fixed number of pixels from each ROI.

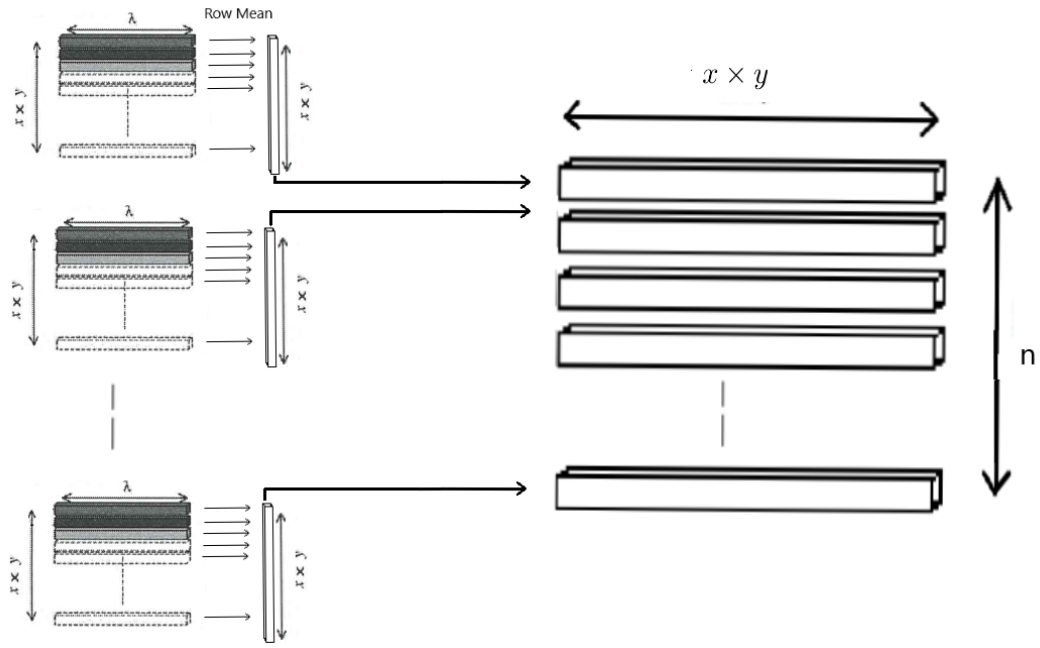


Figure 3.6: Visualisation of the Row-Average data matrix, that is formed by taking the average for each row in the spectral data matrix. Adapted from Roggo et al. (2005)

Compared to the first data matrix in 3.2.2, this matrix has a high computational cost. It also loses information from each individual wavelength and is not able to identify optimal wavelengths. Fig 3.6 visualizes the data extraction process.

3.2.4 Individual Wavelengths

Lastly, the third data format proposed involves using individual wavelengths for each sheep image after selecting and extracting information from the ROI in the hypercube. The resulting format, after combining all images, would be a $m \times 101 \times 31$ datacube, where m is the number of pixels in the ROI.

This datacube is the most computationally expensive of the three proposed data matrices. The first step, as with the row-average matrix, is to sample the number of pixels, which is pre-defined. Then, the first matrix containing only data from the first wavelength for each of the images is used as input to the multivariate method(s) and the results obtained. This is then repeated for each of the remaining wavelengths in the datacube.

The advantage of this data format compared to the other two, is that it will be possible to determine which wavelength is able to most accurately predict the IMF content or classify the marbling score. The drawback, besides computational cost, is that a model containing a combination of information from several wavelengths is not possible.

3.3 Multiple Linear Regression (MLR)

MLR is an extension of simple linear regression models. A simple linear regression aims at explaining the relationship between one continuous response variable and one explanatory variable. However, in most cases there is more than one potential response or explanatory variable, therefore MLR aims at explaining the relationship between the continuous response variable(s) and two or more explanatory variables (Pan et al. (2016); D. Wu, Shi, et al. (2012)). The MLR model equation can be written as:

$$\mathbf{Y} = \beta\mathbf{X} + \epsilon. \tag{3.1}$$

MLR would not be suitable unless there is prior knowledge of which explanatory variables (spectral bands) to include in the model, since depending on the spectral resolution used in acquiring the hyperspectral image, the number of variables (spectral bands) would range from 10 to hundreds. As such other statistical/machine learning techniques are generally preferred.

A common adaptation used in multivariate analysis is penalized linear regression for quantitative regression or penalized logistic regression for qualitative classification (Kyung et al. (2010)). They are also referred to as shrinkage methods, a terminology which we will use henceforth. In particular, the least abso-

lute shrinkage and selection operator (known simply as lasso), ridge regression and elastic net are commonly used penalized regression methods. These adaptations penalize the explanatory variables and minimize the residual sum of squares based on the penalties applied. The penalties are normally applied on the individual regression coefficients, which can cause them to tend to zero, in other words shrink. For the lasso and ridge regression, the loss function that minimizes the residual sum of squares are given below:

$$Loss_{lasso} = (\mathbf{Y} - \mathbf{X}'\beta)'(\mathbf{Y} - \mathbf{X}'\beta) + \lambda|\beta|$$

$$Loss_{ridge} = (\mathbf{Y} - \mathbf{X}'\beta)'(\mathbf{Y} - \mathbf{X}'\beta) + \lambda\beta'\beta$$

The loss function for both the lasso and ridge regression allow explanatory variables (spectral bands) that are weakly related to the response variable(s) to tend to zero, reducing the number of variables included in the final model. The drawback of ridge regression is the final model still incorporates all explanatory variables, though variables that are weakly related to the response variables have a coefficient close to zero. A drawback of the lasso is when two or more variables are highly correlated, the lasso tends to choose one at random so that the final model may include variables that are highly significant in predicting the response but may not make much practical sense. These drawbacks led to the introduction of the elastic net by Zou and Hastie (2005). The elastic net, in simple terms, is a mixture of both the lasso and ridge regression without their inefficiencies.

3.4 Principal Component Analysis (PCA) and Regression (PCR)

PCA is a dimension reduction method developed by Pearson (1901) that converts possible correlated explanatory variables, the wavelengths in our case, into a linear combination of uncorrelated variables called principal components (PC). The first PC aims to explain the greatest amount of variability in the spectral data matrix and the second PC aims at explaining any variability not explained by the first PC and the third PC aiming to explaining any variability not explained by the first two PC and so on. Eventually, the maximum number of PCs would match the number of explanatory variables, where such a case would indicate that all spectral bands are uncorrelated with each other. Technically, PCs are eigenvectors of the covariance matrix of the spectral data matrix with the eigenvalue of the corresponding eigenvector denoting the variance in the data along the eigenvector. PCA then sorts the eigenvectors in decreasing order of their eigenvalue. Principal component regression (PCR) is

an extension of PCA that uses the PCs that have high eigenvalues and forms a MLR using the model (3.1) with the response variables being the object's attributes of interest, such as the IMF content in our case.

3.5 Discriminant Analysis (DA)

DA is a supervised classification method that aims at maximizing the degree of separation between classes (binomial or categorical) in as few dimensions as possible. There are many forms of DA in use with LDA being more widely used than any other. LDA uses discriminant functions, linear combinations of the explanatory variables, that provide the greatest separation between the classes. The first discriminant function provides the greatest separation, after which the separation decreases with each successive discriminant function. The maximum number of discriminant functions is determined by the $min\{p, g-1\}$, with g being the number of classes in the categorical response variable and p being the number of spectral bands in the dataset. LDA bears a similarity to PCA since it also aims at dimension reduction by forming combinations of the explanatory variables, but unlike PCA which is unsupervised and therefore looks at explaining the variance in the spectral data matrix, LDA uses the response variable class and aims at maximizing the separation between the classes (Fisher et al. (1936)).

3.6 Partial Least Squares (PLS)

PLS-R is another dimension reduction method that bears similarity to PCR, developed by Wold et al. (1984). As discussed above, PCR projects the variables onto a new space (PC) that is able to describe as much of the variation in the data as possible but there is no guarantee that the PC with the highest eigenvalue will be relevant to the prediction of the response. PLS-R, on the other hand, is able to use the response variable when choosing the projection onto the new space (called Latent variable (LV)). Therefore, PLS-R chooses projections that would describe the possible covariation between the explanatory variables and the response variable. The projection onto the LVs is a linear combination of the original explanatory variables, with the first LV capturing the biggest variation and correlation between the explanatory variables and the response variables. A regression model is then built using the latent variables. By extending the PLS method to allow use of DA, PLS is capable of performing classification task. This extension is usually denoted as PLS-DA.

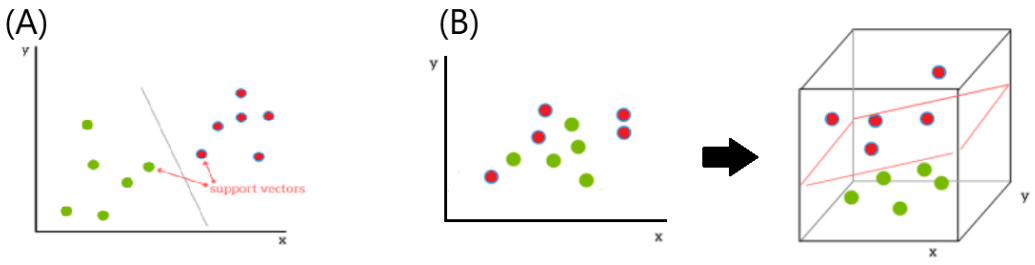


Figure 3.7: Support vector machines illustration using (A) a simple example with clearly separable classes ; (B) Example with non-separable classes

3.7 Support vector machine (SVM)

SVM is a supervised classification algorithm that finds one or more hyperplanes which best divides the data into two or more classes (binary or categorical respectively). To define what a hyperplane is in this context, imagine a simple 2-class example (Fig. 3.7A), i.e. the response variables has only 2 classes (-1, +1), with a line that can clearly separate the data when plotted in a scatterplot. The hyperplane in this example would be the line that linearly separates and classifies the data into the two classes. The position of this line is dependent on the data points nearest to it, these data points are referred to as support vectors. Continuing with the simple 2-class example, finding the optimal hyperplane would require finding the greatest possible margin (distance between the hyperplane and the support vectors) between the hyperplane and the data that would result in a greater chance of any new data being correctly classified (Cortes and Vapnik (1995)).

However, most cases aren't this simple, therefore consider a case where the data can not be clearly linearly separated using a line when plotted in a scatterplot (fig. 3.7B). In this case, SVM maps the data into higher and higher dimensions until a hyperplane can be formed to separate the data into 2 classes. This method of mapping data into higher dimensions is referred to as the kernel trick (or kerneling). Note that when applying the kernel trick, the hyperplane can no longer be a line and therefore becomes a plane (Camps-Valls and Bruzzone (2005)). This forms the basis of the SVM algorithm. Extending the SVM with a least-squares kernel allows for a non-linear regression model to be created that can be used in explaining non-linear data (Suykens and Vandewalle (1999)).

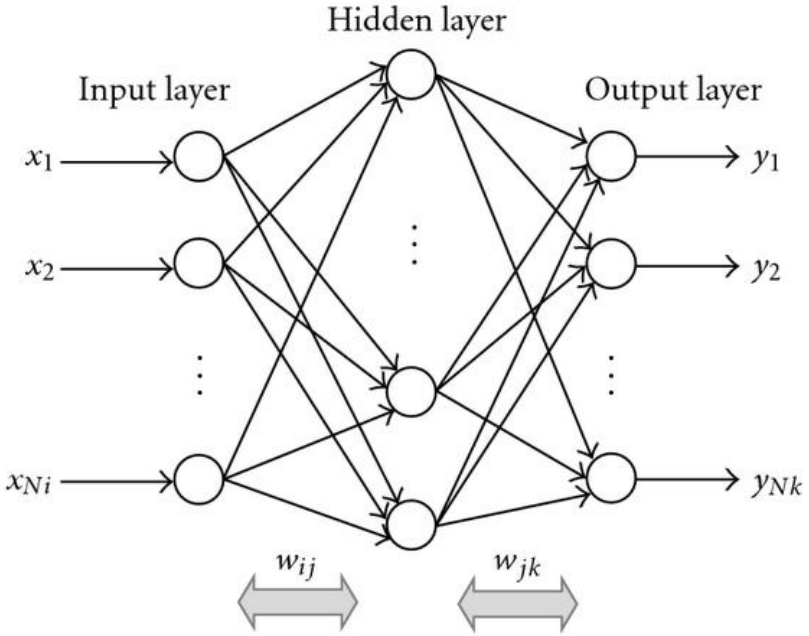


Figure 3.8: Basic structure of a feed-forward Artificial Neural Network

3.8 Artificial Neural Network (ANN)

ANNs are not used in this project, however they are widely used in real world applications and it is therefore worth briefly mentioning. An ANN is an algorithm designed to mimic the organization and knowledge acquisition skills of the human brain. It consists of input and output layers, as well as either one or more hidden layers as illustrated in Fig. 3.8. The layers of an ANN consist of a number of highly interconnected processors, referred to as neurons, that are connected by weighted links passing signals from one neuron to another. The neurons may receive more than one input from the input signals, but output only one signal through the outgoing connection that branches out to the input signals of other neurons in the network. This forms the basic structure and process that an ANN goes through when analysing data for either classification or regression. The most used ANN is the back-propagation neural network, a type of feed-forward neural network. Fig. 3.8 shows a simple feed-forward neural network. In such a network, the input signals are propagated in a forward direction in a layer by layer formation (Lorente et al. (2013)). Thus for a back-propagation neural network, the neural network uses the training data to propagate the input pattern layer by layer until the output is generated. If the generated output is different from the desired output then the error is calculated and propagated backwards through the network from the output layer to the input layer. As the errors gets propagated backwards, the weights are readjusted and modified.

Aside from the great accuracy an ANN provides in most cases, there are two big drawbacks to the method. ANNs are labeled as black boxes because there is no indication as to the formula the ANN used to produce the generated output. ANNs also require a large amount of training data in order to have a high accuracy, which in some cases can be expensive or impossible to acquire. Increasing the number of hidden layers or the number of hidden neurons in the hidden layer can improve the accuracy but this is at the cost of the computational time it takes to train the neural network. Regardless of these drawbacks, ANNs are increasingly widely used.

3.9 Validation and evaluation

To validate the regression or classification models developed in this thesis, the original dataset was randomly split into a 75:25 ratio with the training set having 75% of the dataset and the testing set having 25% of the dataset. The model was developed and calibrated using 10-fold repeated cross-validation with 10 repetitions. That is, the training set of images was further split into 10 folds with approximately the same number of images in each fold. A model was built using 9 of the 10 folds and tested on the remaining fold and the performance metric for the test set calculated. This was repeated 10 times with one model created for each fold. Based on either the minimum Root-mean-square-error (RMSE) or the maximum Accuracy measure, the best model in the training process was chosen and tested using the test images to predict the IMF content or to classify the marbling score and the error calculated. Note that Accuracy, in this case is defined as the number of correctly classified images divided by the total number of images.

RMSE was chosen as the performance measure for the regression techniques as it captures the spread of the residuals, while Accuracy was used for the classification methods to check how well they were able to classify new data.

Chapter 4

Results

In this chapter we look at the results obtained after performing analysis on the pre-processed set of images. The chapter is divided into: Section 4.1, shows how we extracted the sheeps spectral data from the hypercube. In Section 4.2, we present our results of the classification analysis on the sheep data, and in section 4.3, we present our results when using continous IMF content value.

4.1 Spectral data extraction

The white pixels (i.e. the non-zero elements) in the final mask (Fig. 2.6) of the sheep images after the masking step was used as the ROI to extract the sheep carcass (spectral data) from the hyperspectral image. Then, we employed the column-average data matrix discussed in section 3.2.2, that is, the dataset was formed by calculating the average spectral value of all pixels in the ROI for each wavelength. This procedure was then repeated to obtain the average value from all 101 sheeps in our dataset and saved in a matrix $\mathbf{X}$ with the sheeps as rows and the wavelengths as the columns. The matrix was then pre-processed using mean-centering before any subsequent analysis was conducted.

Fig 4.1 shows the distribution of the mean reflectance values in the ROI per wavelength. We can see that there appears to be several outliers, with the most significant being sheep images 151, 160 and 163. As there is no reason to delete the values, we will conduct two analyses with one being the full data set (named *full*, n = 101) and one where the potential outliers (named *reduced*, n = 98) were removed (see Appendix Fig. 5.2).

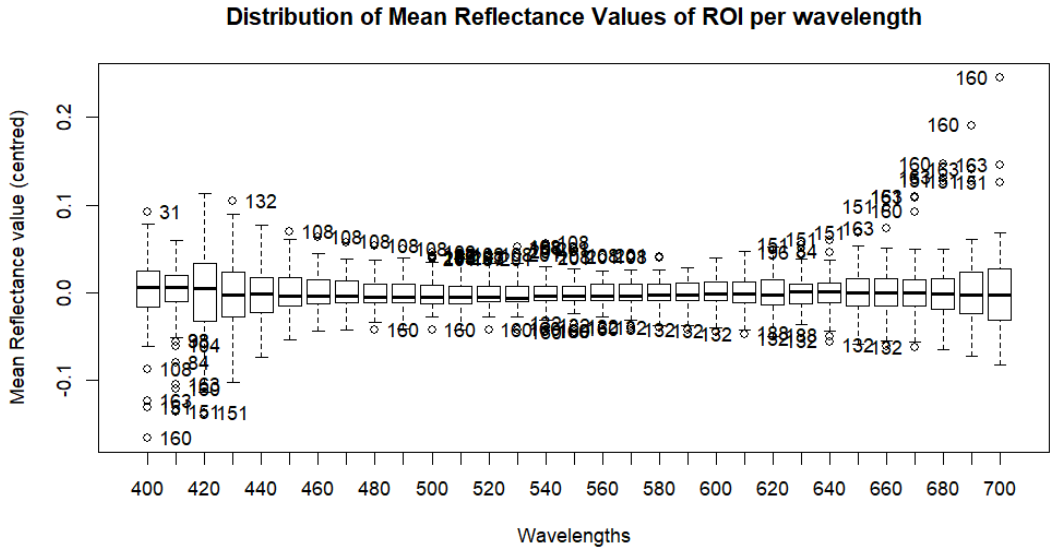


Figure 4.1: Boxplot of the Sheep carcasess showing the distribution of the mean reflectance values of the ROI per wavelength

4.2 Classifying Marbling Score

The classification models used to classify the marbling score were briefly introduced in section 3.3, specifically LDA, PLS-DA, SVM and MLR using the logistic regression. As mentioned previously, LDA and PLS-DA both perform dimension reduction by either using discriminant functions that provide the greatest separation between marbling scores or by capturing the biggest variation and correlation between wavelengths and marbling scores. In contrast, SVMs find one or more hyperplanes which best divides the data into two or more marbling scores and has the ability to be kerneled allowing non-linear data to be analysed. The MLR aims to explain the relationship between the binary reponse (e.g. the marbling score) and the wavelengths, where use of shrinkage methods allows for dimension reduction.

In this section, we look at finding the optimal parameters using the training set of images. Afterwards, we test the calibrated model using the test set and compare the performance of the different techniques.

4.2.1 Tuning Parameters

Besides the traditional MLR technique, all other methods consist of parameters that define them. For LDA and PLS-DA, the number of discriminant functions and latent variables respectively, needs to be determined. For SVMs, the cost parameter, that controls the trade-off between a hyperplane with the largest minimum margin and a hyperplane that correctly separates as many instances

as possible, needs to be determined. While for the shrinkage methods, the regularization terms (λ) needs to be determined.

Fig. 4.2 shows the results of the parameter tuning using 10-fold repeated cross-validation as described in section 3.9 on the *full* dataset. Fig. 4.2 (a) shows the number of discriminant functions against model accuracy for the LDA method. We can see that there is no difference between them, therefore, the number of discriminant functions chosen for the LDA model is 1 with an accuracy of 0.59. This is expected as LDA chooses the discriminant function using $min\{p, g-1\}$, where $p = 31$ and $g = 2$. In Fig. 4.2 (b), it is clear that the number of latent variables with the highest accuracy for the PLS-DA function is 21 with an accuracy of 0.61. For the SVM (c), the cost parameter that had the highest accuracy in the training set was 0.3, with an accuracy of 0.545. The ridge regression method (d) had an accuracy of approximately 0.54 with a regularization value of 0.1. The lasso method (e) had a regularization value of 0.0205 with an accuracy of approximately 0.512. Lastly, the elastic net had an accuracy of 0.538 for regularization values of 0.1 and 0.03.

For the *reduced* dataset (see Appendix Fig. 5.3), the parameters for LDA and SVM for both datasets were similar to the *full* dataset, however the accuracy for the LDA model was less by approximately 0.07, while the accuracy for the SVM was greater by approximately 0.04. There was a slight change in the parameters for the shrinkage methods, with the lasso regularization parameter being slightly less (0.0161) for the *reduced* dataset, however had an increased accuracy by 0.05. The ridge model had a regularization value of 0.16, slightly more than the same model using the *full* dataset, with an accuracy of 0.56 which is approximately similar to the same model using the *full* dataset. The elastic net model parameters for the two dataset were similar, with an accuracy of approximately 0.54 for both datasets. Lastly, the PLS-DA model had 3 latent variables for the *reduced* dataset, a difference of 18 from the same model using the *full* dataset. There was not much difference in the accuracy of the models using both datasets.

4.2.2 Results

With the models developed and calibrated, we used the test set of sheep images to determine whether the models are able to correctly classify the marbling score of the 'new' hyperspectral sheep images. Table 4.1 compares the accuracy measurements of the models using the test images.

It can be seen that the shrinkage methods performed better in classifying the marbling score using the test set of images of the *reduced* dataset, with accuracy

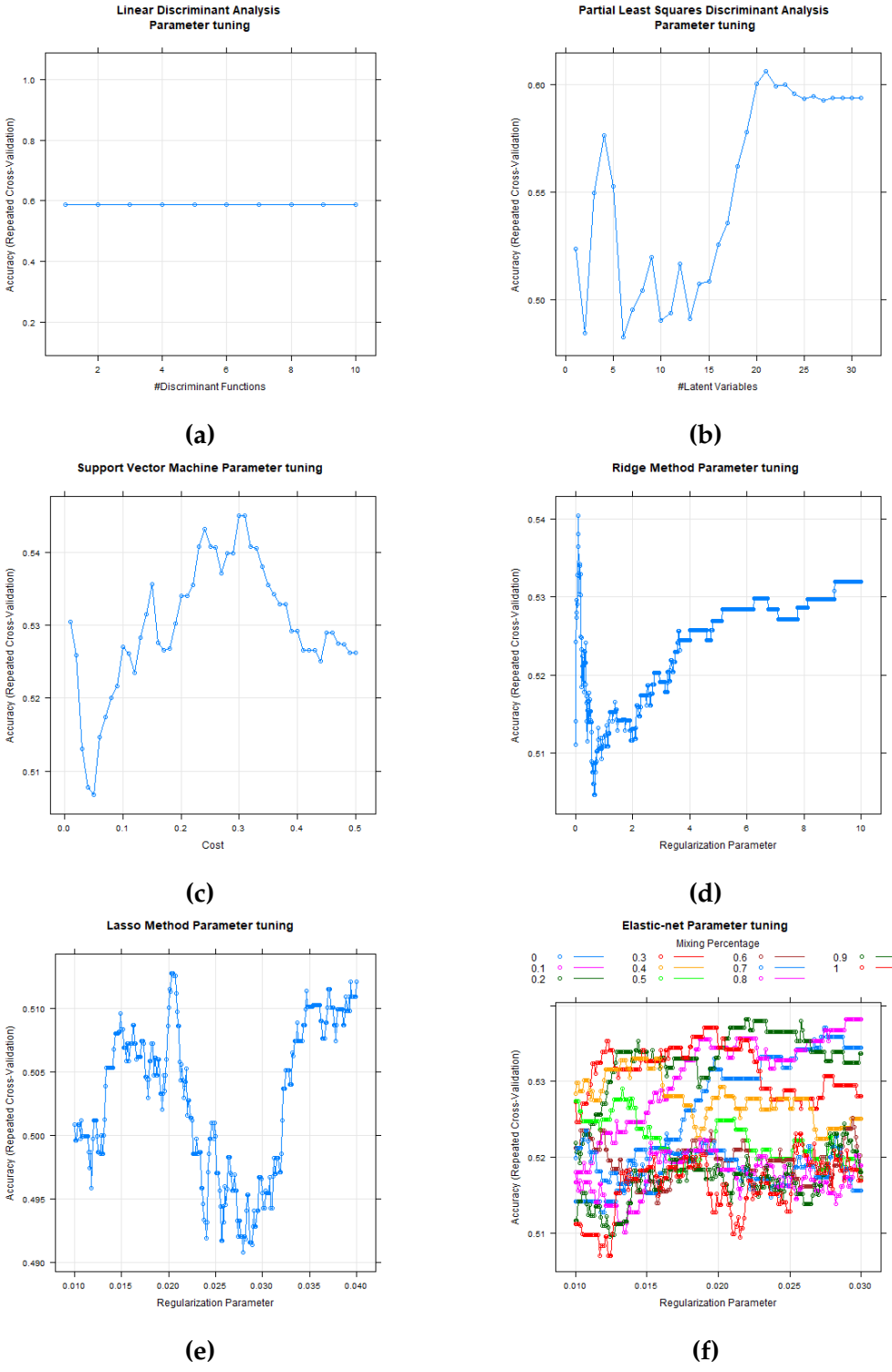


Figure 4.2: Parameter tuning using 10-fold cross-validation on the *full* dataset for (a) LDA (b) PLS-DA (c) SVM (d) ridge method (e) lasso method (f) elastic net method

Table 4.1: Classifaction model performance metrics for the test images for both the *full* and *reduced* datasets. The specificity and sensitivity is also recorded.

Classifaction Model Metrics

Models	***full***		***reduced***	
	$Specificity\ (Sens)$	$Accuracy$	$Specificity\ (Sens)$	$Accuracy$
MLR	0.27 (0.69)	0.50	0.45 (0.42)	0.43
Lasso	0.15 (0.77)	0.50	0.45 (0.67)	0.57
Ridge	0.18 (0.69)	0.46	0.45 (0.75)	0.60
Elastic-net	0.36 (0.62)	0.54	0.45 (0.75)	0.61
LDA	0.45 (0.46)	0.46	0.36 (0.58)	0.49
PLS-DA	0.45 (0.38)	0.42	0.45 (0.58)	0.52
SVM	0.36 (0.77)	0.54	0.45 (0.67)	0.52

values of 0.61, 0.60 and 0.57 for the elastic net, ridge and lasso models respectively. The LDA and MLR had accuracies of 0.49 and 0.43 respectively, on the *reduced* dataset. And the PLS-DA and SVM models had an accuracy of 0.52, on the *reduced* dataset.

For the *full* dataset, the shrinkage methods had accuracies of 0.54, 0.50 and 0.46 for the elastic net, lasso and ridge models respectively. These were less than the accuracies observed when using the *reduced* dataset, suggesting that the models were affected by the outliers. In fact this seem to be the case for the LDA and PLS-DA, with both models having lower accuracies when using the *full* dataset than the *reduced* dataset. The MLR and SVM are the only models that seem not to be affected by the outliers, with both having an increase in accuracy of 0.07 and 0.02 respectively.

4.2.3 Model Specific Wavelength Importance

Using the final model developed during the training phase, we are able to determine the importance of a specific wavelength in a model. The importance measurement measures the percentage of "influence" a wavelength has on the model. For linear models, the percentage is measured by the absolute value of the t-statistic for each model parameter. For the shrinkage methods, the absolute value of the wavelength coeffiecients corresponding to the final

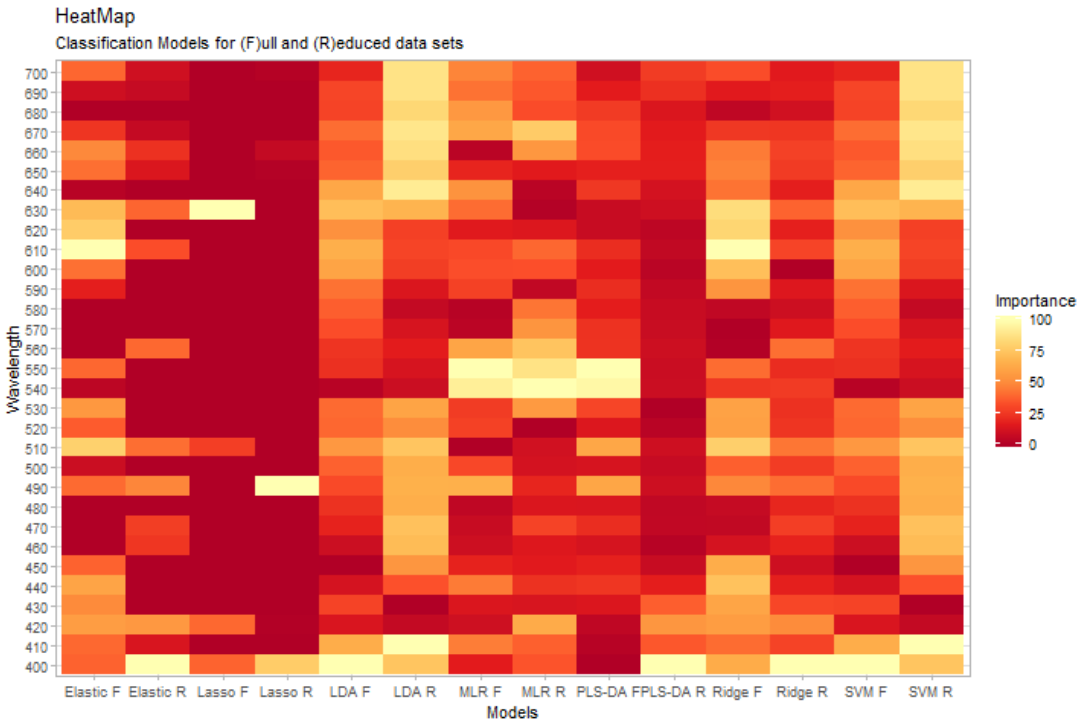


Figure 4.3: Importance percentage of the spectral bands shown as a Heatmap for the Classification models. Where the *full* dataset is denoted by **F** and the *reduced* dataset is denoted by **R**

model are used. For the LDA, PLS and PCR, the measurement is based on the weighted sums of the absolute regression coeffiecents. And for the SVM, a loess smoother is fit between the marbling score and the wavelength and the R^2 statistic is then calculated for the model against the intercept only model.

Fig. 4.3 shows the heatmap of the importance percentage of the spectral bands against the classification models. From this, we see that the lasso models for both data sets have the least number of spectral bands in the final model. This is expected given the property of the lasso to shrink non-important variables to zero. In the lasso model for the *reduced* data, only 3 spectral bands 490nm, 400nm and 660nm were important. The least important band in the final lasso model, 660nm, has an importance percentage of less than 25%. The LDA and SVM models appear to have similar spectral bands for the *full* and *reduced* data set. The *reduced* data set for both models has at least 20 bands with importance percentages of above 50%, while PLS-DA, elastic net and lasso models had only about 3-4 bands above 50%.

The 400nm spectral band appears in all classification model at varying levels of importance. The 420nm band appears in all but the lasso model when considering the *reduced* dataset while the 490nm band appears in all models except the lasso with the *full* data set. The 630nm band also appears in all models ex-

cept the lasso and the MLR model with the *reduced* data set, as does the 660nm band in the *full* dataset.

The bands with 100% importance value in different models were: 610nm for the elastic net and ridge method on the *full* data set, 400nm for the elastic net, PLS-R and ridge method on the *reduced* data set and the LDA on the *full* data set, 630nm for the lasso on the *full* data set, 490nm for the lasso on the *reduced* data set, 410nm for the LDA and the SVM on the *reduced* dataset, 550nm for PLS-DA and MCR models on the *full* dataset, 540nm for the MCR on the *reduced* data set. We also note that the bands 570nm and 580nm appear to have an average importance of less than 15% for all models, with both not appearing in either the elastic net or lasso model for both data sets under consideration.

4.3 Predicting Percentage of Intramuscular Fat Content

To predict the IMF content, we use regression techniques introduced in section 3.3. These are PCR, PLS-R, SVM, MLR and shrinkage methods. As mentioned previously, PCR is a dimension reduction technique that uses a linear combinatination of principal components as explanatory variables. The other methods used in this section have already been briefly mentioned in section 4.2.

In this section, we look at finding the optimal parameters using the training set of images. We then test the calibrated model using the test set of images and compare the different techniques.

4.3.1 Tuning Parameters

The parameter tuning process for the regression techniques is similar to that of the classification techniques in section 4.2.1. The difference being that the outcome variable is continous instead of binary, i.e. the IMF content instead of the marbling score. The other difference is the selection of the best model. Instead of maximizing the Accuracy of the model as was done in the classification case, here we are minimizing the RMSE. The parameters that require tuning for the PLS-R, SVM and MLR with shrinkage methods have been mentioned in section 4.2.1. For PCR, the number of principal components needs to be determined.

Fig. 4.4 shows the results of the parameter tuning using 10-fold repeated cross-validation as described in section 3.9 on the *full* dataset. Fig. 4.4 (a) shows that

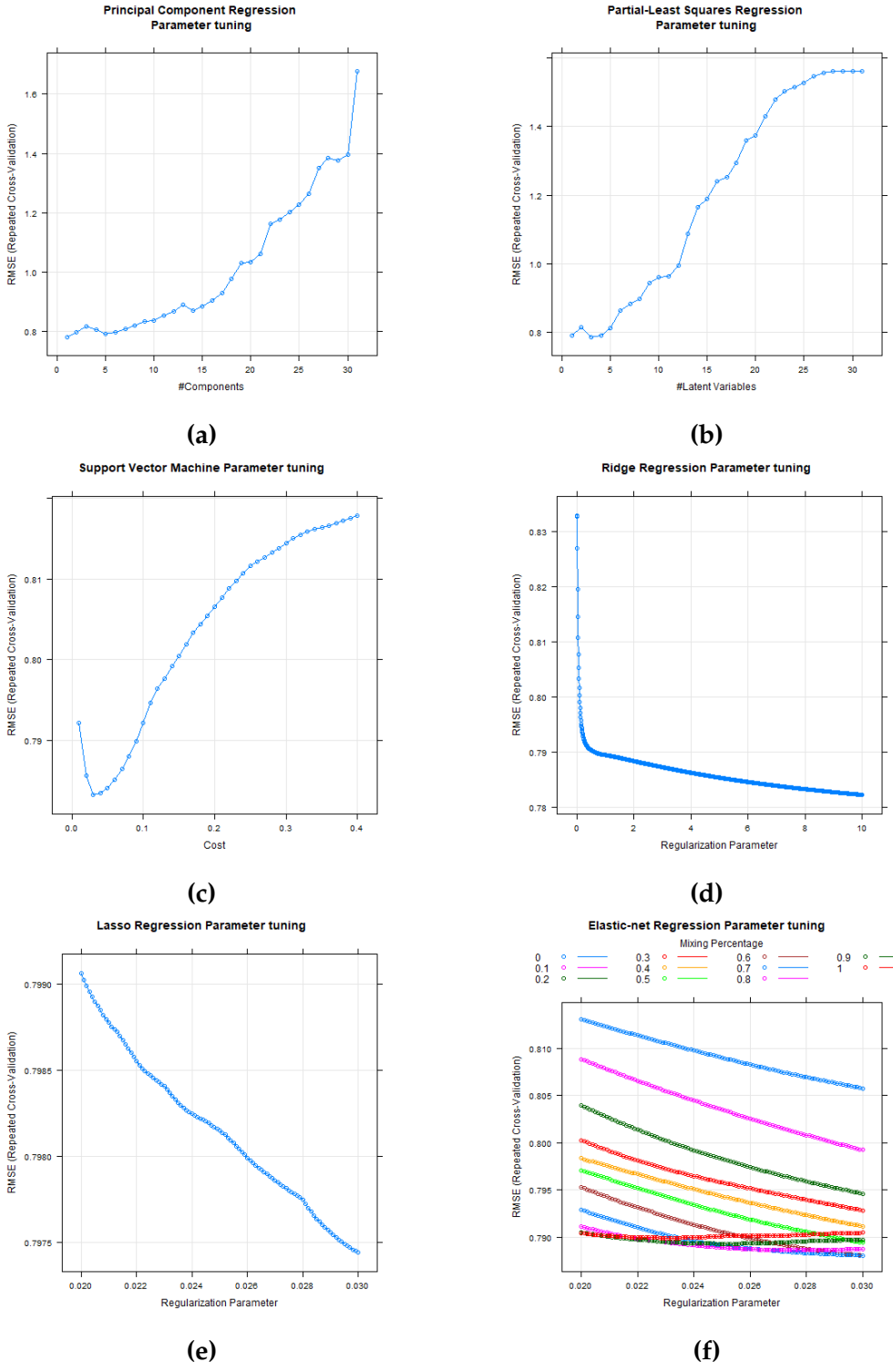


Figure 4.4: Parameter tuning using 10-fold cross-validation on the *full* dataset for (a) PCR (b) PLS-R (c) SVM (d) ridge method (e) lasso method (f) elastic net method. *Full* dataset is denoted by **F** and *reduced* dataset by **R**

the number of principal components against the model RMSE for the PCR method. We can see the number of principal components chosen for the PCR model is 1 with a RMSE value of 0.779. In Fig. 4.4 (b), it is clear that the number of latent variables with the lowest RMSE for the PLS-R method is 3 with a RMSE value of 0.786. And for the SVM (c), the cost parameter that had the lowest RMSE value in the training set was 0.03, with RMSE value of 0.783. Ridge regression (d) had a RMSE value of approximately 0.782 with a regularization value of 10. The lasso method (e) had a regularization value of 0.03 with a RMSE value of approximately 0.797. Lastly, the elastic net had a RMSE value of 0.788 for the regularization values of 0.7 and 0.03.

For the *reduced* dataset see Appendix Fig. 5.4. The parameters for two of the shrinkage methods, i.e. lasso and ridge, had the same value for the parameter for both datasets under consideration. The RMSE was slightly higher in the *full* dataset than the *reduced* dataset for both models (0.721 and 0.707 respectively). PLS-R and PCR had 2 latent variables and 3 principal components, respectively, on the *reduced* dataset with an RMSE value of approximately 0.693, which were the best models in the training phase for both datasets. The SVM had a cost parameter of 0.01 on the *reduced* dataset with an RMSE difference of 0.07 between the two datasets. Lastly, the elastic net had regularization parameters of 1 and 0.03 with a RMSE of 0.715, a difference of 0.073 from the *full* dataset. Overall, the *reduced* dataset had better performance in the training phase with an average RMSE difference of 0.157 between the two datasets

4.3.2 Results

Once the models were developed and calibrated, we used the test set of sheep images to measure the RMSE of the models. That is, we measured how far the predicted IMF content values were from the actual IMF content values. Table 4.2 compares the RMSE measurements of the models using the *full* and *reduced* datasets.

It can be seen that the shrinkage methods performed better in predicting the IMF content using the test set of images of the *full* dataset, with RMSE values of 0.583, 0.595 and 0.597 for the lasso, elastic net and ridge models respectively. The PCR and MLR had RMSE values of 0.601 and 0.958 respectively, on the *full* dataset. And the PLS-R and SVM models had RMSE values of 0.606 and 0.609 respectively, on the *full* dataset.

For the *reduced* dataset, the shrinkage methods had RMSE values of 0.894 for both elastic net and lasso models, while the RMSE value of the ridge model

Table 4.2: Regression model performance metrics for the test images for both the *full* and *reduced* datasets. The R^2 and Mean-Absolute-Error (MAE) are also recorded.

Regression Model Metrics

Models	***full***			***reduced***		
	RMSE	R^2	*MAE*	*RMSE*	R^2	*MAE*
MLR	0.958	0.00436	0.759	1.572	0.199	1.133
Lasso	0.583	0.0632	0.492	0.894	0.00204	0.686
Ridge	0.597	0.0671	0.492	0.901	0.105	0.688
Elastic-net	0.595	0.0525	0.504	0.894	0.00204	0.686
PCR	0.601	0.0139	0.494	0.893	0.000793	0.676
PLS-R	0.606	0.0519	0.514	0.960	0.0759	0.717
SVM	0.609	0.0513	0.499	0.974	0.1089	0.724

was 0.901. These were greater than the RMSE values observed when using the *full* dataset, suggesting that the models were affected by the outliers. In fact this seem to be the case for the PCR and PLS-R, with both models having higher RMSE values when using the *reduced* dataset than the *full* dataset. Similarly for the MLR and SVM models, with RMSE values of 1.572 and 0.974.

Clearly, the results presented in Table 4.2 suggest that the outliers had an effect in the prediction of IMF content on the test set of sheep images, with models for the *full* dataset performing better than for the *reduced* dataset. The best model identified for the *reduced* dataset was the PCR model, although the lasso and elastic net model had almost similar performance (difference of 0.001 in RMSE value). For the *full* dataset, the best model was the lasso for predicting the IMF content in the test set of sheep images.

4.3.3 Model Specific Wavelength Importance

Using the final model developed during the training phase, we are able to determine the importance (see section 4.2.3) of a specific wavelength on a model. Fig. 4.5 shows the heatmap of the importance percentage of the spectral bands against each of the regression models.

From the figure, we see that the lasso and elastic net models appear to be sim-

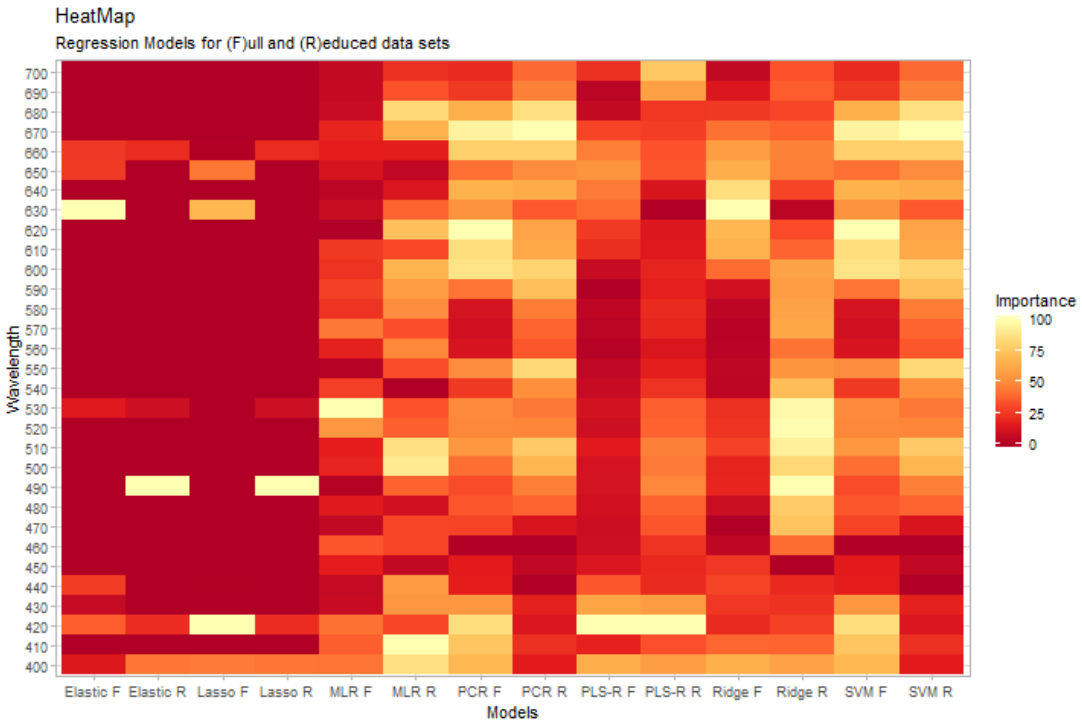


Figure 4.5: Importance percentage of the spectral bands shown as a Heatmap for the Regression models. Where the *full* dataset is denoted by **F** and the *reduced* dataset is denoted by **R**

ilar as determined by the same number of wavelengths deemed important for the *reduced* dataset. This is expected as both models had similar regularization values as seen in Fig. 4.4. We also note that both the elastic net and lasso regression for both data set have the smallest number of spectral bands in the final model after the calibration of the model using the training set. While the PCR, PLS-R and SVM models for the *full* dataset had almost every band, besides the 460nm band for the PCR and SVM models and 630nm for the PLS-R model, in the final model.

It can be noted that the spectral bands 400nm and 420nm appear in all regression models at varying levels of importance. The spectral bands 530nm and 660nm also appear in every regression model, besides the lasso on the *full* dataset. The bands deemed very important in the models were: 630nm for the elastic net and ridge regression on the *full* data set, 490nm for the elastic net, lasso and ridge regression on the *reduced* data set, 420nm for the lasso on the *full* data set and PLS-R for both data sets, 530nm for the MLR on the *full* data set, 410nm for the MLR on the *reduced* dataset, 620nm for the PCR and SVM on the *full* data set and 680nm for the PCR and SVM on the *reduced* data set. Lastly, we note that the bands 450nm and 460nm appear to have an average importance of approximately 10% for all models, with both not appearing in the elastic net and lasso regression for both data sets.

Chapter 5

Discussion

The use of hyperspectral imaging analysis for the prediction and classification of food quality and food safety has proven to be a powerful analytical tool for non-destructive analysis (Gowen et al. (2007)). This thesis has demonstrated the application of this technique in predicting the IMF content and classifying the marbling score of sheep carcasses. Prior to the analysis, reflectance calibration was conducted using the methods described in section 2.3. Image segmentation to differentiate between the sheep pixels and the background was conducted using k-means. The sheep pixels were then extracted into three datasets discussed in section 3.2 to be used as inputs to classification and regression methods discussed in section 3.3. Several regression and classification models were created, calibrated and validated using two datasets, *full* and *reduced*. The *reduced* dataset did not contain data from 3 sheep images (150, 160 and 161) as they appeared to be extreme values with average pixel values exceeding the interquartile range for each wavelength in the dataset (see Fig. 4.1). These extreme value were removed as they may affect the development and prediction accuracy of the models, however models were created using the *full* dataset and the results of the two datasets compared.

As seen in Table 4.2, we found that the models performed better in predicting new hyperspectral image data using the *full* dataset. The results obtained are comparable to the results reported by National ICT Australia (NICTA) for determining IMF content in sheep carcasses (Robles-Kelly and R. Wei (2015)). We compared the Mean-Absolute-Error of our models and that of the model created by NICTA (= *0.748*), and found that almost all our models performed better except the MLR model. This can be attributed to the fact that, while we opted to use K-means to segment and extract our ROI in the spatial pre-processing stage, they used all pixels in the image segmented using a method they developed which is akin to K-means (Robles-Kelly and R.

Wei (2015)).

We also developed and validated models that would classify the hyperspectral images into a categorical marbling score, based on IMF content. Using the Accuracy metric to evaluate a models performance, we found in this instance, performance of the models was slightly better if the *reduced* dataset was used. This was expected as classification methods are highly susceptible to the presence of extreme values, affecting the model's efforts to generalize. The elastic net model was identified as the best model with an accuaracy of 61%, indicating that the model would have a 61% probability of correctly classifying a new image. Overall the classification model performances were not significantly different to random chance unfortunately.

In spite of being able to predict the IMF content and classify the marbling score using hyperspectral images, we were unable to test the process using the row-average and individual wavelength data formats. We found that the maximum number of pixels that could be used as independent variables was approximately 15,000 that accounted to 3-7.5% of the pixels in one sheep image. Any more variables would cause memory issues, as the resulting data matrix would be too big to hold in memory and perform analysis on it. This also posed the problem of overfitting in the development of the models in the training phase, as there are more independent variables than the number of sheep images. We also note that the Modest and the Slight– group had to be merged with the Small and Slight+ groups respectively due to the small number of sheep in the dataset. We suspect, this loss of information could have attributed to the poor performance of the classification models. In fact, we found the LDA model had a better classification accuracy (results not shown) when only considering dataset with Small and Slight+ marbling scores.

Despite these limitations, we were able to showcase the predictive ability of the hyperspectral image analysis. In this thesis, we were able to show the effectiveness of hyperspectral image analysis in the prediction of IMF content from hyperspectral sheep images. Possible future works includes consideration to using more robust multivariate techniques to reduce the effects outliers would have in the model calibration and validation. We may also consider using high performance computing in the analysis given the vast amount of processing power it may have, especially in the case for the row-average and individual wavelength where the number of independent variables exceed 15,000. We may also consider a further segmentation of the masked image to different parts such as the muscle, bone, fat etc. and performing the analysis on a combination of the average values of the newly segmented areas for each wavelength to predict the IMF content. Lastly, as the analysis was conducted in the visible

spectrum, perhaps performing the same analysis in the near-infrared spectrum would yield better prediction accuracy.

Appendix A

Spearman Correlation

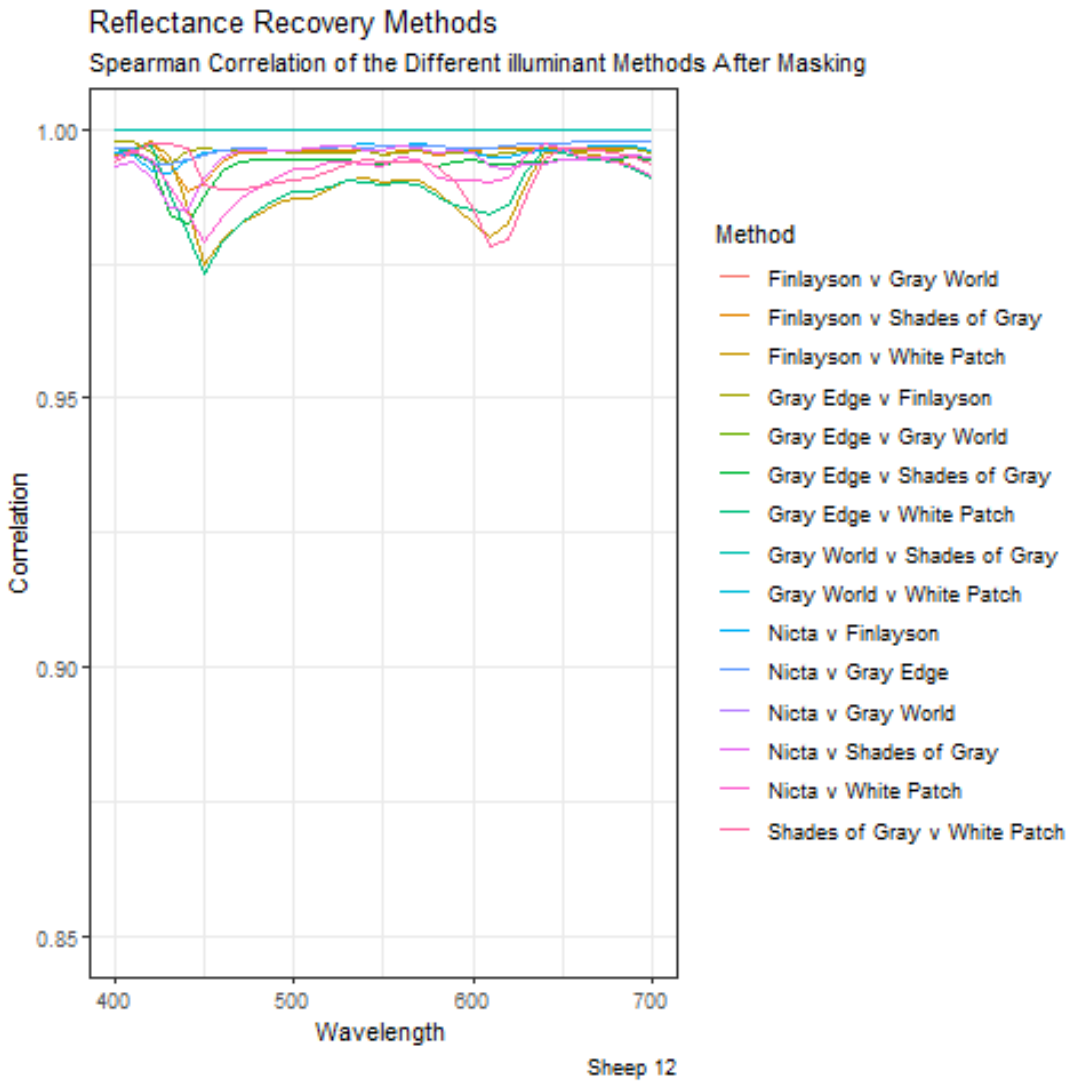


Figure 5.1: Spearman Correlation graph for all possible combinations of the 6 different methods Grey World, White Patch, NICTA Reflectance, Shades of Grey, Grey Edge and Finlayson and Shaefer

Appendix B

Extreme Values

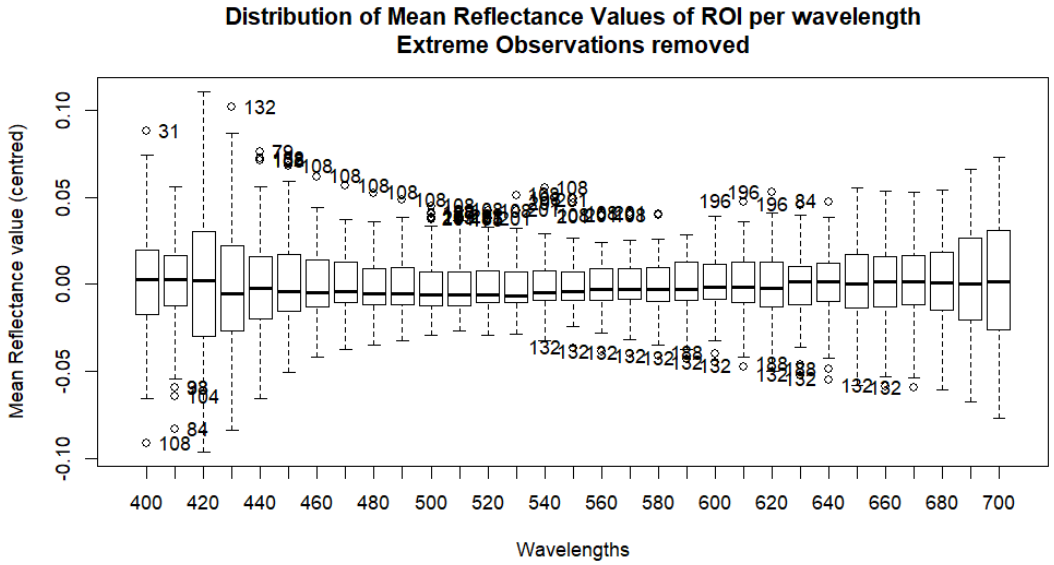


Figure 5.2: Boxplot of the mean reflectance value for each wavelength with outliers removed

Appendix C

Parameter Tuning

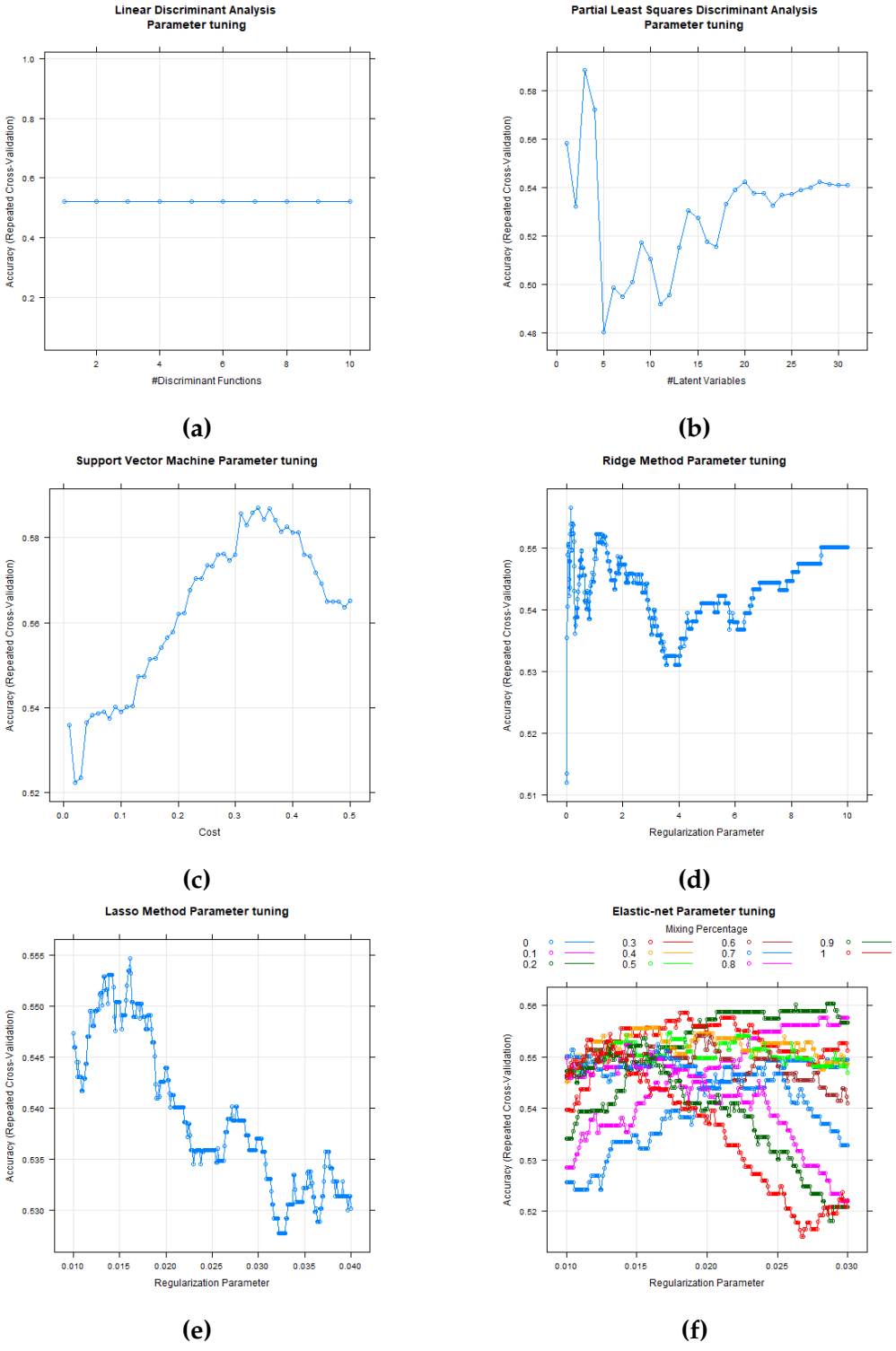


Figure 5.3: Parameter tuning using 10-fold cross-validation on the *reduced* dataset for (a) LDA (b) PLS-DA (c) SVM (d) ridge method (e) lasso method(f) elastic net method

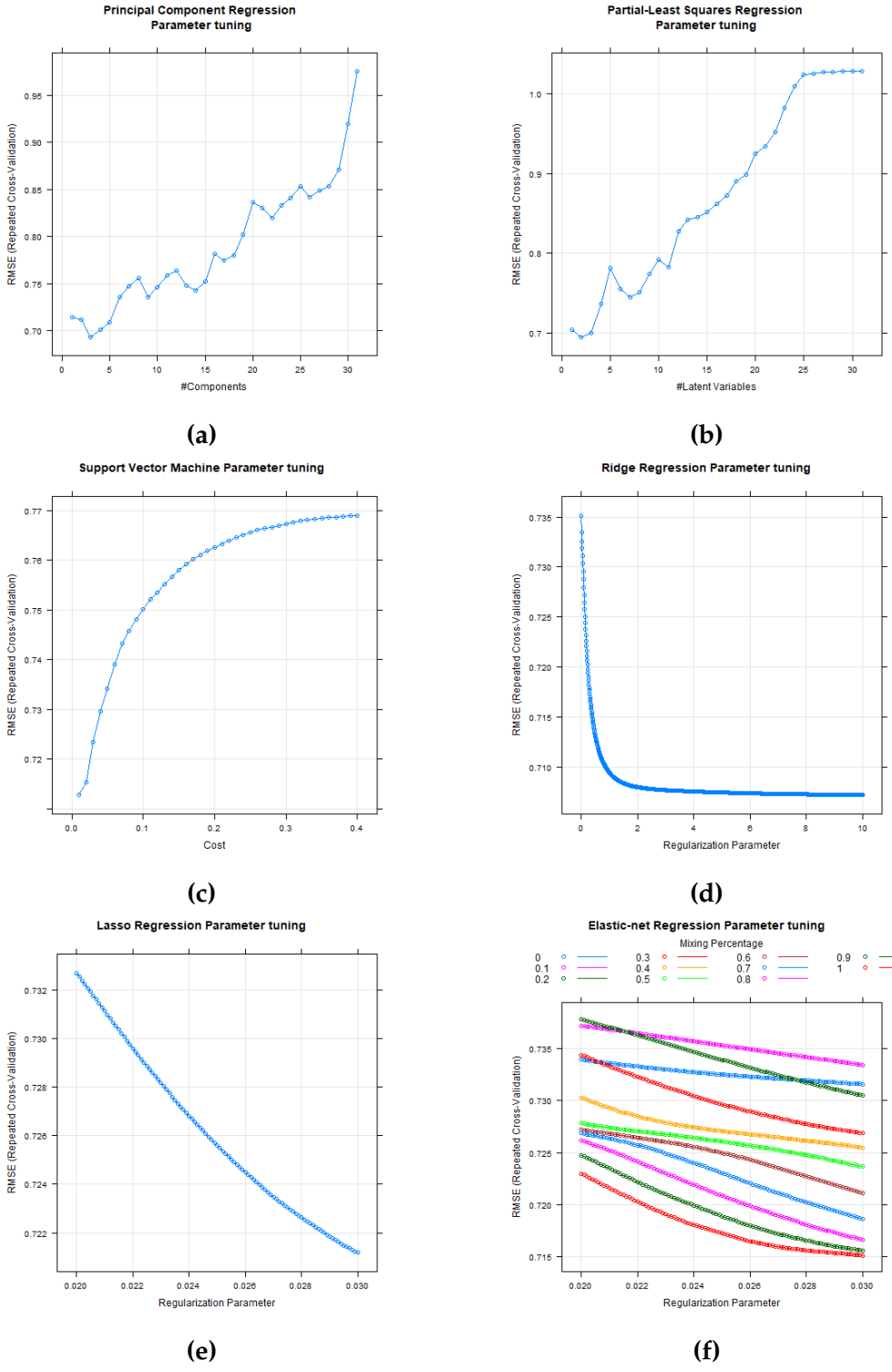


Figure 5.4: Parameter tuning using 10-fold cross-validation on the *reduced* dataset for (a) PCR (b) PLS-R (c) SVM (d) ridge regression (e) lasso regression (f) elastic net regression

Bibliography

Abdullah, MZ et al. (2004). "The applications of computer vision system and tomographic radar imaging for assessing physical properties of food". In: *Journal of food engineering* 61(1), pp. 125–135.

Adam, Elhadi, Onisimo Mutanga, and Denis Rugege (2010). "Multispectral and hyperspectral remote sensing for identification and mapping of wetland vegetation: a review". In: *Wetlands Ecology and Management* 18(3), pp. 281–296.

Adão, Telmo et al. (2017). "Hyperspectral imaging: A review on UAV-based sensors, data processing and applications for agriculture and forestry". In: *Remote Sensing* 9(11), p. 1110.

Amigo, José Manuel, Hamid Babamoradi, and Saioa Elcoroaristizabal (2015). "Hyperspectral image analysis. A tutorial". In: *Analytica Chimica Acta* 896, pp. 34–51.

Al-Amri, Salem Saleh, Namdeo V Kalyankar, et al. (2010). "Image segmentation by using threshold techniques". In: *arXiv preprint arXiv:1005.4020*.

Bock, JE and RK Connelly (2008). "Innovative uses of near-infrared spectroscopy in food processing". In: *Journal of Food Science* 73(7), R91–R98.

Boesch, Chris et al. (1997). "In vivo determination of intra-myocellular lipids in human muscle by means of localized 1H-MR-spectroscopy". In: *Magnetic resonance in medicine* 37(4), pp. 484–493.

Bong, Chin-Wei and Mandava Rajeswari (2011). "Multi-objective nature-inspired clustering and classification techniques for image segmentation". In: *Applied soft computing* 11(4), pp. 3271–3282.

Buchsbaum, Gershon (1980). "A spatial processor model for object colour perception". In: *Journal of the Franklin institute* 310(1), pp. 1–26.

Burnase, Miss Sneha R and Miss Shanti Swamy (2002). "Hyperspectral image reduction using discrete wavelet transform". In: *Remote Sensing* 40(10).

Calin, Mihaela Antonina et al. (2015). "Characterization of burns using hyperspectral imaging technique–A preliminary study". In: *Burns* 41(1), pp. 118–124.

Camps-Valls, Gustavo and Lorenzo Bruzzone (2005). "Kernel-based methods for hyperspectral image classification". In: *IEEE Transactions on Geoscience and Remote Sensing* 43(6), pp. 1351–1362.

Cen, Haiyan and Yong He (2007). "Theory and application of near infrared reflectance spectroscopy in determination of food quality". In: *Trends in Food Science & Technology* 18(2), pp. 72–83.

Chen, Chenyi et al. (2015). "Deepdriving: Learning affordance for direct perception in autonomous driving". In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2722–2730.

Cloutis, Edward A (1996). "Review article hyperspectral geological remote sensing: evaluation of analytical techniques". In: *International Journal of Remote Sensing* 17(12), pp. 2215–2242.

Cortes, Corinna and Vladimir Vapnik (1995). "Support-vector networks". In: *Machine learning* 20(3), pp. 273–297.

Du, Cheng-Jin and Da-Wen Sun (2006). "Learning techniques used in computer vision for food quality evaluation: a review". In: *Journal of food engineering* 72(1), pp. 39–55.

ElMasry, Gamal M and Shigeki Nakauchi (2016). "Image analysis operations applied to hyperspectral images for non-invasive sensing of food quality–A comprehensive review". In: *Biosystems engineering* 142, pp. 53–82.

ElMasry, Gamal, Da-Wen Sun, and Paul Allen (2013). "Chemical-free assessment and mapping of major constituents in beef using hyperspectral imaging". In: *Journal of Food Engineering* 117(2), pp. 235–246.

ElMasry, Gamal, Ning Wang, and Clément Vigneault (2009). "Detecting chilling injury in Red Delicious apple using hyperspectral imaging and neural networks". In: *Postharvest biology and technology* 52(1), pp. 1–8.

Esquerre, C et al. (2012). "Suppressing sample morphology effects in near infrared spectral imaging using chemometric data pre-treatments". In: *Chemometrics and Intelligent Laboratory Systems* 117, pp. 129–137.

Feng, Chao-Hui et al. (2018). "Hyperspectral imaging and multispectral imaging as the novel techniques for detecting defects in raw and processed meat products: Current state-of-the-art research advances". In: *Food Control* 84, pp. 165–176.

Finlayson, Graham D and Gerald Schaefer (2001). "Convex and non-convex illuminant constraints for dichromatic colour constancy". In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*. Vol. 1. IEEE, pp. I–I.

Finlayson, Graham D and Elisabetta Trezzi (2004). "Shades of gray and colour constancy". In: *Color and Imaging Conference*. Vol. 2004. 1. Society for Imaging Science and Technology, pp. 37–41.

Fisher, Ronald Aylmer et al. (1936). “138: The Use of Multiple Measurements in Taxonomic Problems.” In:

Gowen, AA et al. (2007). “Hyperspectral imaging–an emerging process analytical tool for food quality and safety control”. In: *Trends in food science & technology* 18(12), pp. 590–598.

Habili, Nariman and Jeremy Oorloff (2015). “Scyllarus^{TM}: From Research to Commercial Software”. In: *Proceedings of the ASWEC 2015 24th Australasian Software Engineering Conference*. ACM, pp. 119–122.

Hadsell, Raia et al. (2009). “Learning long-range vision for autonomous off-road driving”. In: *Journal of Field Robotics* 26(2), pp. 120–144.

Halicek, Martin et al. (2018). “Tumor margin classification of head and neck cancer using hyperspectral imaging and convolutional neural networks”. In: *Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling*. Vol. 10576. International Society for Optics and Photonics, p. 1057605.

Hassen, Abebe et al. (2001). “Predicting percentage of intramuscular fat using two types of real-time ultrasound equipment”. In: *Journal of animal science* 79(1), pp. 11–18.

He, Hong-Ju and Da-Wen Sun (2015). “Selection of informative spectral wavelength for evaluating and visualising Enterobacteriaceae contamination of salmon flesh”. In: *Food analytical methods* 8(10), pp. 2427–2436.

Hocquette, JF et al. (2010). “Intramuscular fat content in meat-producing animals: development, genetic and nutritional control, and identification of putative markers”. In: *Animal* 4(2), pp. 303–319.

Huang, Hui, Li Liu, and Michael O Ngadi (2014). “Recent developments in hyperspectral imaging for assessment of food quality and safety”. In: *Sensors* 14(4), pp. 7248–7276.

Huynh, Cong Phuoc and Antonio Robles-Kelly (2008). “Optimal solution of the dichromatic model for multispectral photometric invariance”. In: *Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR)*. Springer, pp. 382–391.

Huynh, Cong Phuoc and Antonio Robles-Kelly (2010). “A solution of the dichromatic model for multispectral photometric invariance”. In: *International Journal of Computer Vision* 90(1), pp. 1–27.

Kamruzzaman, Mohammed, Gamal ElMasry, et al. (2012a). “Non-destructive prediction and visualization of chemical composition in lamb meat using NIR hyperspectral imaging and multivariate regression”. In: *Innovative Food Science & Emerging Technologies* 16, pp. 218–226.

Kamruzzaman, Mohammed, Gamal ElMasry, et al. (2012b). “Prediction of some quality attributes of lamb meat using near-infrared hyperspectral imaging and multivariate analysis”. In: *Analytica Chimica Acta* 714, pp. 57–67.

Kamruzzaman, Mohammed, Gamal ElMasry, et al. (2013). “Non-destructive assessment of instrumental and sensory tenderness of lamb meat using NIR hyperspectral imaging”. In: *Food Chemistry* 141(1), pp. 389–396.

Kamruzzaman, Mohammed, Yoshio Makino, and Seiichi Oshita (2015). “Non-invasive analytical technology for the detection of contamination, adulteration, and authenticity of meat, poultry, and fish: a review”. In: *Analytica chimica acta* 853, pp. 19–29.

Kamruzzaman, Mohammed, Yoshio Makino, and Seiichi Oshita (2016). “Parsimonious model development for real-time monitoring of moisture in red meat using hyperspectral imaging”. In: *Food chemistry* 196, pp. 1084–1091.

Keresztes, Janos C et al. (2017). “Glare based apple sorting and iterative algorithm for bruise region detection using shortwave infrared hyperspectral imaging”. In: *Postharvest biology and technology* 130, pp. 103–115.

Kuhn, Max (2008). “Building Predictive Models in R Using the caret Package”. In: *Journal of Statistical Software, Articles* 28(5), pp. 1–26. ISSN: 1548-7660. DOI: 10.18637/jss.v028.i05. URL: https://www.jstatsoft.org/v028/i05.

Kyung, Minjung et al. (2010). “Penalized regression, standard errors, and Bayesian lassos”. In: *Bayesian Analysis* 5(2), pp. 369–411.

Leavesley, Silas J et al. (2016). “Hyperspectral imaging fluorescence excitation scanning for colon cancer detection”. In: *Journal of biomedical optics* 21(10), p. 104003.

Lee, Craig A et al. (2011). “Recent developments in high performance computing for remote sensing: A review”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4(3), pp. 508–527.

Lee, Woo-Ram, Dong-Guk Hwang, and Byoung-Min Jun (2011). “Comparison of color constancy methods for skin color under colored illuminants”. In: *The 7th International Conference on Digital Content, Multimedia Technology and its Applications*. IEEE, pp. 80–83.

Li, Qingli et al. (2013). “Review of spectral imaging technology in biomedical engineering: achievements and challenges”. In: *Journal of biomedical optics* 18(10), p. 100901.

Lohumi, Santosh et al. (2016). “Application of hyperspectral imaging for characterization of intramuscular fat distribution in beef”. In: *Infrared Physics & Technology* 74, pp. 1–10.

Lonergan, Steven M, David G Topel, and Dennis N Marple (2018). *The Science of Animal Growth and Meat Technology*. Academic Press.

Lorente, Delia et al. (2013). "Selection of optimal wavelength features for decay detection in citrus fruit using the ROC curve and neural networks". In: *Food and Bioprocess Technology* 6(2), pp. 530–541.

Lu, Guolan and Baowei Fei (2014). "Medical hyperspectral imaging: a review". In: *Journal of biomedical optics* 19(1), p. 010901.

Mager, Donald E and Darrell R Abernethy (2007). "Use of wavelet and fast Fourier transforms in pharmacodynamics". In: *Journal of Pharmacology and Experimental Therapeutics* 321(2), pp. 423–430.

Mehl, Patrick M et al. (2004). "Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations". In: *Journal of Food Engineering* 61(1), pp. 67–81.

Miclos, Sorin et al. (2015). "Algorithm for mapping cutaneous tissue oxygen concentration using hyperspectral imaging". In: *Biomedical optics express* 6(9), pp. 3420–3430.

Naganathan, Govindarajan Konda et al. (2016). "Three dimensional chemometric analyses of hyperspectral images for beef tenderness forecasting". In: *Journal of food engineering* 169, pp. 309–320.

Nanyam, Yasasvy et al. (2012). "A decision-fusion strategy for fruit quality inspection using hyperspectral imaging". In: *Biosystems engineering* 111(1), pp. 118–125.

Newcom, DW, TJ Baas, and JF Lampe (2002). "Prediction of intramuscular fat percentage in live swine using real-time ultrasound". In: *Journal of Animal Science* 80(12), pp. 3046–3052.

Ng, Hui-Fuang (2006). "Automatic thresholding for defect detection". In: *Pattern recognition letters* 27(14), pp. 1644–1649.

Nicolai, Bart M et al. (2007). "Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review". In: *Postharvest biology and technology* 46(2), pp. 99–118.

Nunes, Jose Luis et al. (2015). "Beef quality parameters estimation using ultrasound and color images". In: *BMC bioinformatics* 16(4), S6.

Oliveira, Leandro S and Adriana S Franca (2011). "Applications of Near Infrafred Spectroscopy (NIRS) in food quality evaluation". In: *Food Quality: Control, Analysis and Consumer Concerns*, pp. 131–179.

Pan, Ting-Tiao et al. (2016). "Regression algorithms in hyperspectral data analysis for meat quality detection and evaluation". In: *Comprehensive Reviews in Food Science and Food Safety* 15(3), pp. 529–541.

Pavurala, Naresh, Xiaoming Xu, and Yellela SR Krishnaiah (2017). "Hyperspectral imaging using near infrared spectroscopy to monitor coat

thickness uniformity in the manufacture of a transdermal drug delivery system". In: *International journal of pharmaceutics* 523(1), pp. 281–290.

Pearson, Karl (1901). "LIII. On lines and planes of closest fit to systems of points in space". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 2(11), pp. 559–572.

Pohi, C and JL Van Genderen (1998). "Multisensor image fusion in remote sensing: concepts, methods and applications". In: *I Remote Sensing* 19(5), pp. 823–854.

Prakash, Bhaskaran David and Yap Chun Wei (2011). "A fully automated iterative moving averaging (AIMA) technique for baseline correction". In: *Analyst* 136(15), pp. 3130–3135.

Pu, Hongbin, Mohammed Kamruzzaman, and Da-Wen Sun (2015). "Selection of feature wavelengths for developing multispectral imaging systems for quality, safety and authenticity of muscle foods-a review". In: *Trends in Food Science & Technology* 45(1), pp. 86–104.

Pu, Yuan-Yuan, Yao-Ze Feng, and Da-Wen Sun (2015). "Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: a review". In: *Comprehensive Reviews in Food Science and Food Safety* 14(2), pp. 176–188.

Rahman, Sejuti and Antonio Robles-Kelly (2013). "An optimisation approach to the recovery of reflection parameters from a single hyperspectral image". In: *Computer vision and image understanding* 117(12), pp. 1672–1688.

Ravn, Carsten, Erik Skibsted, and Rasmus Bro (2008). "Near-infrared chemical imaging (NIR-CI) on pharmaceutical solid dosage forms—comparing common calibration approaches". In: *Journal of Pharmaceutical and Biomedical Analysis* 48(3), pp. 554–561.

Rinnan, Åsmund, Frans Van Den Berg, and Søren Balling Engelsen (2009). "Review of the most common pre-processing techniques for near-infrared spectra". In: *TrAC Trends in Analytical Chemistry* 28(10), pp. 1201–1222.

Robles-Kelly, Antonio and Ran Wei (2015). "Project Report for the Phase 2A (Validating Predictions of Intra-muscular Fat Content) of the Project Hyperspectral lmaging for Grading Meat from Sheep Carcasses". In: p. 15.

Roggo, Y et al. (2005). "Infrared hyperspectral imaging for qualitative analysis of pharmaceutical solid forms". In: *Analytica Chimica Acta* 535(1-2), pp. 79–87.

Sabins, Floyd F (1999). "Remote sensing for mineral exploration". In: *Ore Geology Reviews* 14(3-4), pp. 157–183.

Samopa, Febriliyan and Akira Asano (2009). "Hybrid image thresholding method using edge detection". In: *International Journal of Computer Science and Network Security* 9(4), pp. 292–299.

Savitzky, Abraham and Marcel JE Golay (1964). "Smoothing and differentiation of data by simplified least squares procedures." In: *Analytical chemistry* 36(8), pp. 1627–1639.

Shafer, Steven A (1985). "Using color to separate reflection components". In: *Color Research & Application* 10(4), pp. 210–218.

Sun, Da-Wen and Cheng-Jin Du (2004). "Segmentation of complex food images by stick growing and merging algorithm". In: *Journal of Food Engineering* 61(1), pp. 17–26.

Suykens, Johan AK and Joos Vandewalle (1999). "Least squares support vector machine classifiers". In: *Neural processing letters* 9(3), pp. 293–300.

Van De Weijer, Joost, Theo Gevers, and Arjan Gijsenij (2007). "Edge-based color constancy". In: *IEEE Transactions on image processing* 16(9), pp. 2207–2214.

Van der Meer, Freek D et al. (2012). "Multi-and hyperspectral geologic remote sensing: A review". In: *International Journal of Applied Earth Observation and Geoinformation* 14(1), pp. 112–128.

Wahabzada, Mirwaes et al. (2017). "Monitoring wound healing in a 3D wound model by hyperspectral imaging and efficient clustering". In: *PloS one* 12(12), e0186425.

Wilczyński, Sławomir et al. (2016). "The use of hyperspectral imaging in the VNIR (400–1000 nm) and SWIR range (1000–2500 nm) for detecting counterfeit drugs with identical API composition". In: *Talanta* 160, pp. 1–8.

Wold, Svante et al. (1984). "The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses". In: *SIAM Journal on Scientific and Statistical Computing* 5(3), pp. 735–743.

Wu, Di, Hui Shi, et al. (2012). "Rapid prediction of moisture content of dehydrated prawns using online hyperspectral imaging system". In: *Analytica Chimica Acta* 726, pp. 57–66.

Wu, Di and Da-Wen Sun (2013a). "Advanced applications of hyperspectral imaging technology for food quality and safety analysis and assessment: A review—Part I: Fundamentals". In: *Innovative Food Science & Emerging Technologies* 19, pp. 1–14.

Wu, Di and Da-Wen Sun (2013b). "Colour measurements by computer vision for food quality control–A review". In: *Trends in Food Science & Technology* 29(1), pp. 5–20.

Wu, I-Chen et al. (2018). "Early identification of esophageal squamous neoplasm by hyperspectral endoscopic imaging". In: *Scientific reports* 8(1), p. 13797.

Yang, Guang et al. (2007). "Study on statistics iterative thresholding segmentation based on aviation image". In: *Eighth ACIS International*

Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing (SNPD 2007). Vol. 2. IEEE, pp. 187–188.

Zeaiter, M, J-M Roger, and V Bellon-Maurel (2005). "Robustness of models developed by multivariate calibration. Part II: The influence of pre-processing methods". In: *TrAC Trends in Analytical Chemistry* 24(5), pp. 437–445.

Zhang, Yu Jin (1996). "A survey on evaluation methods for image segmentation". In: *Pattern recognition* 29(8), pp. 1335–1346.

Zheng, Chaoxin, Da-Wen Sun, and Liyun Zheng (2006). "Recent developments and applications of image features for food quality evaluation and inspection–a review". In: *Trends in Food Science & Technology* 17(12), pp. 642–655.

Zou, Hui and Trevor Hastie (2005). "Regularization and variable selection via the elastic net". In: *Journal of the royal statistical society: series B (statistical methodology)* 67(2), pp. 301–320.

Zuzak, Karel J et al. (2001). "Noninvasive determination of spatially resolved and time-resolved tissue perfusion in humans during nitric oxide inhibition and inhalation by use of a visible-reflectance hyperspectral imaging technique". In: *Circulation* 104(24), pp. 2905–2910.