QUALITY ESTIMATION OF CANOLA USING MACHINE VISION AND VIS-NIR SPECTROSCOPY

By

GEETIKA DILAWARI

Bachelor of Technology in Agricultural Engineering
Punjab Agricultural University
Ludhiana, Punjab, India
2004

Master of Science in Agricultural and Biosystems Engineering
Iowa State University
Ames, IA
2006

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2011

QUALITY ESTIMATION OF CANOLA USING MACHINE VISION AND VIS-NIR SPECTROSCOPY

Dissertation Approved:

Dr. Carol Jones

Dissertation Adviser

Dr Randy Taylor

Dr Marvin Stone

Dr Neils Maness

Outside Committee Member

Dr. Sheryl A. Tucker

Dean of the Graduate College

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

Canola belongs to the Brassicaceae (Cruciferae) family and is a Canadi an invention. Though, Canola was bred from rapeseed its fatty acid profile is different from that of rapeseed. It has high amounts of monounsaturated fatty acids (oleic acid, C18:1, 60%), moderate levels of essential polyunsaturated fatty acids, linoleic acid (C18:2, 20%) and α-linolenic acid (C18:3, 10%) and very low levels of erucic acid (C 22:1, <2%). Canola seeds contain around 44 percent oil. It is considered as one of the healthiest oils due to low amount of saturated fatty acids and is becoming popular among people due to its nutritive value. High amounts of unsaturated fatty acids reduce cholesterol levels, hence canola oil is considered good for heart patients. Canola oil was granted Generally Recognized as Safe (GRAS) status for human consumption in 1985 by United States (U.S.) Food and Drug Administration (Berglund et al 2007). Canola oil is also considered as a potential source for manufacturing biodiesel. In addition to high nutritive value of canola oil, canola meal is a good source of protein for animals. It contains around 38% protein. In the U.S, canola is mainly grown in North Dakota (ND), Oklahoma (OK), Minnesota, Montana, Idaho and Oregon. Total current acreage of canola in the U.S. is around 1.5 million acres with supply to demand ratio of canola oil around 1:3(U.S.

Canola Association). North Dakota produces around 90 percent of domestic production and Oklahoma is the second largest producer of canola (NASS 2010). In 2010 total production of the U.S. was around 2.45 billion pounds out of which ND produced around 2.18 billion pounds and OK around 89.6 million pounds (NASS 2011). In northern parts of the U.S. and Canada, canola is grown in spring where as in OK it is grown in winter. The yield of winter canola is considered to be 20-30% more than the spring canola (Boyles, Canola Project Specialist, OKANOLA, Oklahoma State University, Accessed on March 2011).

The grade determination of canola seeds in the U.S. mainly follows the U.S. standard guidelines (USDA). According to U.S. standards, canola is graded as U.S. No.1, 2, 3, and sample grade. Table 1.1 shows the grade requirements of canola.

Table 1.1 U.S Standard Guidelines for Grading of Canola

Grade	Maximum Limits of							
	Damaged Kernels			Conspicuous Admixture				Inconspicuous Admixture (%)
	Heat Damaged (%)	Distinctly Green (%)	Total (%)	Ergot (%)	Sclerotinia (%)	Stones (%)	Total (%)	
U.S No. 1	0.1	2.0	3.0	0.05	0.05	0.05	1.0	5.0
U.S No. 2	0.5	6.0	10.0	0.05	0.1	0.05	1.5	5.0
U.S No. 3	2.0	20.0	20.0	0.05	0.15	0.05	2.0	5.0

U.S. Sample Grade

(a) Does not meet the requirements for grades U.S. No. 1,2,3; or

(b) Contains 1 or more pieces of glass, 2 or more particles of an unknown foreign substance (s) or a commonly recognized harmful or toxic substance (s), or 4 or more pieces of animal filth;

(c) Has a musty , sour, or commercially objectionable foreign odor; or

(d) Is heating or otherwise of distinctly low quality.

Green seeds, damaged seeds, conspicuous admixture, broken seeds, contaminated grain, animal excreta, foreign material, inconspicuous admixture, insect excreta, and stones mainly affect canola quality grading (Table 1.1). The grading of seeds is mainly done by visual inspection. Rancidity causes musty or sour odor of the seeds as a result, decreases the grade of canola to US sample grade. U.S. graders primarily smell a sample of seeds to grade canola on the basis of odor. But these odors can be easily eliminated or suppressed by a producer by using fumigants or other odor suppressants. These odors are mainly present due to the presence of ketones and aldehydes as the oil in seeds undergoes oxidation (Caddick, SGRL) or in other words when the oil goes rancid. Rancidity is mainly caused by hydrolysis of triglycerides to free fatty acids (FFA) and then oxidation of FFA into ketones and aldehydes. It can be categorized into hydrolytic rancidity and oxidative rancidity. Presence of high amount of FFA can be used as an indicator of hydrolytic rancidity and oxidative rancidity is measured by determining the peroxide value (PV). PV measures the extent of oxidation by determining the amount of peroxides present. Therefore, percent FFA content or PV can be used as indicators to determine the seed rancidity. According to Canadian General Standards Board requirements for canola oil, FFA content in crude oil should not be more than 1% (by mass as oleic acid). In refined, bleached and deodorized oil the maximum permissible FFA content and PV is 0.05% and 2 meq /kg respectively (Canola: Standards and Regulations, CGSB 1987). The U.S. has not published any similar food grade specifications in regard to canola oil.

U.S. grade defines canola as seeds of Brassica family, which have erucic acid less than 2 percent. Other seeds of Brassica family like rapeseed and mustard have more erucic acid

(around 40-60% of total fatty acids) present in them. These seeds look similar to canola but differ in their fatty acid composition. The popularity of canola is mainly due to its fatty acid composition (high oleic and low erucic acid) therefore, rapeseed and mustard are considered as impurities in canola if present. Other than visual inspection there is no other current method to distinguish between these seeds. Suspected samples are sent for laboratory testing where wet chemical tests estimate the amount of erucic acid. This is a time consuming process and involves destruction of sample.

As discussed above canola is mainly graded either by visual inspection or by smelling. These methods are subjective in nature and are bound to cause errors while deciding the grade of canola. To test canola for amount of erucic acid present the sample needs to be sent to a laboratory for testing through wet chemical analysis. This is a time consuming process. An electronic method that can quantify amount of dockage, presence of distinctly green and heat treated seeds, distinguish samples on the basis of erucic acid, FFA content and PV, would not only be less time consuming but also would be a more reliable method to grade canola samples.

Therefore following were the main objectives outlined for this study:

- Quantification of dockage in canola using flatbed scanner
- Identification of distinctly green and heat damaged seeds in canola using machine vision
- Non destructive classification of canola on the basis of erucic acid using NIR spectroscopy
- Non destructive estimation of free fatty acid content and peroxide value in canola using NIR spectroscopy

CHAPTER II

QUANTIFICATION OF DOCKAGE IN CANOLA USING FLATBED SCANNER

Abstract: Various machine vision techniques have been applied to grade, size and classify different grain types such as wheat, rice, lentils, pulses and soybeans. Little work has been done to grade canola using machine vision. Presence of dockage and conspicuous admixture affects the quality of canola. Therefore, classification of canola samples on the basis of total dockage (machine separated dockage + conspicuous admixture) using flatbed scanners has been outlined as the main objective for this study. Samples with varying amount of total dockage, 0.1, 0.3, 1, 1.34, 1.45, 2.5, 3.4, 4.3, 4.6 and 11.3% were prepared. Classification of samples involved recording the intensity values of Luminosity (L), Red (R), Green (G) and Blue (B) domains of sample images and analyzing the mean sample values by discriminant analysis using PROC GLM and PROC DISCRIM procedures in SAS. The analysis showed a misclassification rate of 14%. The sample with 2.5% foreign material showed maximum error during classification of samples. Sample with maximum amount of total dockage (around 11.3%) was significantly different from all other samples.

Review of Literature

Many researchers have been working on developing electronic techniques to grade different grain types. Machine vision, image spectroscopy and near infrared (NIR)

spectroscopy are among a few of the techniques that are being used for this purpose. Machine vision involving identification and separation of grains by digital image processing is one of the most common techniques being used. Impurities are separated from the grain using different physical features like size, shape, texture, color and other morphological parameters. Morphological and textural parameters have been used to identify the quality of rice by classifying the sample into healthy and defective grains (Bal *et al*, 2006). The quality of rice was evaluated by measuring its length, breadth, area, perimeter, texture and color. High-resolution images, obtained using three chip charge coupled device (CCD) color cameras, have been successfully used to identify different grain kernels. This method uses an algorithm based on kernel signature which involves shape, length and color of different grain kernels. Though this algorithm was able to identify different grain kernels, classification of damaged kernels, foreign material content was not tested (Paliwal *et al*, 1999). Image analysis in conjunction with a back propagation neural network as a classifier has been used to identify different grain types on the basis of color and textural features (Visen *et al*, 2003). Another study on quantification of foreign material (barley) in wheat has shown that back propagation neural networks with statistical classifiers can be used to classify wheat and barley admixture correctly. The neural network showed better performance than the statistical classifier. A need to improve the algorithm so as to improve the efficiency of the classifier was identified. The classification accuracy was less (1.2%) for barley admixture (Tahir *et al*, 2006). Mean accuracy of 100% was observed when bulk samples of Canada Western Red Spring (CWRS) wheat, Canada Western Amber Durum (CWAD) wheat, barley, oats, and rye were classified on the basis of textural features extracted from red

color band (Majumdar and Jayas 1999). Optimization of textural and color features showing significant contribution were identified using PROC STEPDISC and the data was analyzed using PROC DISCRIM procedures in SAS statistical software. Images were acquired using a 3 chip CCD color camera and a 10-120 mm focal length zoom lens. Variance derived from gray-level co-occurrence matrix (GCLM) and hue was considered as the most significant textural and color features, respectively. Majumdar and Jayas (2000) developed another model using 10 most significant morphological features to classify CWRS wheat, CWAD wheat, barley, oats and rye with classification accuracy of 98.9, 91.6, 97.9, 100.0 and 91.6% respectively. Models based on morphological features and combination of morphology-color features showed mean accuracies of 99 and 96.3%, respectively, on training dataset. Mean accuracies of around 90 and 89.4 % for corresponding models was achieved on test data sets when CWRS wheat, durum wheat, barley, rye and oats were classified from their dockage components (Nair, 1997). Machine vision has also been used to quantify the percentage of dockage material in the grain sample before and after it has passed through the cleaner and thus has been used to test the performance of a grain cleaner (Paliwal, 2004). In another study conducted on rapeseed it was observed that color was not a good indicator to distinguish between mature, immature and broken seeds. R, G and B values were able to significantly distinguish between rape and stickywilly seeds (Tanska et al. 2005). One of the studies have concluded that model developed by combination of morphological, color and textural features would show better classification performance (Paliwal et al. 2003) . A model developed by combining morphological and color features showed an improved identification accuracy when machine vision was used to classify healthy and six types of

damaged kernels (Luo *et al*, 1999). Machine vision system has also been used to identify corn kernel mechanical and mold damage with 99.5 and 98.7% accuracy respectively (NG et al. 1998). Another study on soybeans yielded a classification accuracy of 94% when seeds were classified into asymptomatic, immature and discoloured soybean seeds conforming to USDA/FGIS grading procedures (Casady *et al*, 1992). Most of the discussed studies have used the CCD camera as their image acquisition device.

CCD cameras yield high resolution images but are quite expensive. A rather inexpensive machine vision system that has been used and tested by researchers is the flatbed scanner (Paliwal *et al*, 2004; Shahin and Symons, 2001; Shahin *et al*, 2006). Flatbed scanners with back propagation neural networks as a classifier have been successfully used to classify cereal grains using color, textural and morphological features of the samples (Paliwal *et al*, 2004). Machine vision techniques using flat bed scanners have also been applied for determining the seed size distribution of lentil seeds, seed sizing of pulses, color and size grading of pulse grains, seed size uniformity of soybean seeds, and quality of rice (Shahin and Symons, 2001; Shahin *et al*, 2004; Shahin *et al*, 2006; Kumar and Bal, 2006). Thresholding based on histogram analysis in conjunction with other morphological features has also been used to segment damaged seeds (Shatadal and Tan, 2003) in soybeans.

Researchers have also used NIR reflectance spectroscopy for classification of damaged seeds in wheat and soybean. A study has shown the possibility of using NIR spectrum to classify damaged soybean seeds from sound seeds but has also suggested more data collection and further analysis before using this technology commercially (Wang, Ram and Dowell, 2002). The results from this study had demonstrated 99% accuracy when a

two-class model, sound and damaged seeds, was used. The accuracy decreased when damaged seeds were classified into further classes like heat, sprout, frost, mold, and weather-damaged seeds to develop a six-class model. Similar results were observed when damaged wheat kernels, scab and other mold damage were separated from sound wheat kernels. Classification accuracy between 95% and 98% was achieved when wheat kernels were classified as sound or damaged kernels (Delwiche, 2003).

Different techniques have been applied for rice, wheat, pulses, soybeans and lentils to distinguish between good and bad seeds in addition to quantifying the amount of dockage present, but little work has been done to apply these techniques in canola. The main objective of this study is to develop a sensor system for quality grading of the canola samples. This sensor system would be based on the reflectance properties of canola as these would be affected by the amount of total dockage present in the sample. Therefore, data was collected for color features only using a flatbed scanner and the results were validated using the reflectance data obtained from a spectrometer. Classification of canola on the basis of total dockage (machine separated dockage + conspicuous admixture) using a flatbed scanner was outlined as the objective for this study.

Material and Methods

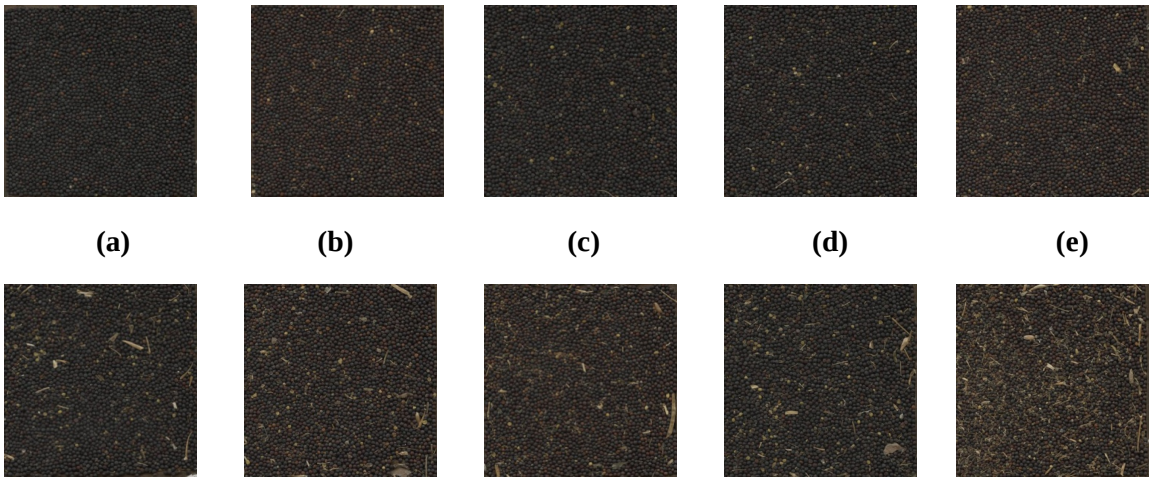
Grain Samples

Canola samples were prepared with variable amount of dockage and conspicuous admixture. These samples were then graded by graders at Enid Grain Inspection Company Inc., Enid, OK. As per the reports from the graders, samples had 0.1, 0.3, 1, 1.34, 1.45, 2.5, 3.4, 4.3, 4.6 and 11.3% total dockage present in them (Figure 2.1). The

dockage and conspicuous admixture included straw, pieces of wood, dead grass, ergot, other grains like wheat and material which is readily distinguishable from canola like stones, insects etc. The weight of each sample was 1000 gm. To account for the variability within each sample, twenty independent sub samples weighing 45gm from each sample were used for further testing.

Image and Data Acquisition

Samples were scanned using a color image flat bed scanner (Epson Perfection V500 photo scanner, Epson America Inc., Long Beach, CA, USA). A wooden frame, 0.127 m x 0.127 m x 0.002 m , was used to hold each sample in a uniform distribution while on the scanner (Figure 2.2 a). For each sample a 512 by 512 pixel image was captured at 150 dpi (dots per inch). Color Checker white balance card (X-rite Incorporation, Grand Rapids, Michigan, USA) was used for color calibration of the scanner. Color calibration was done at the start of image acquisition and thereafter every five images. The correction was applied only if the variation in the reflectance values between the standards ranged above 1%. The mean values, that is the average intensity values, of the luminous (L), red (R), green (G) and blue (B) domains were recorded using Adobe Photoshop 2.0, an image editing software. The RGB model assigns each pixel an intensity value ranging between 0 (black) and 255 (white) for each of the RGB and luminosity components of a color image (Adobe Photoshop Elements 2.0). This represents the visible spectrum. The data for all 20 samples were recorded and averaged to give a mean value for L, R, G and B. These averaged values were then used for further analysis.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

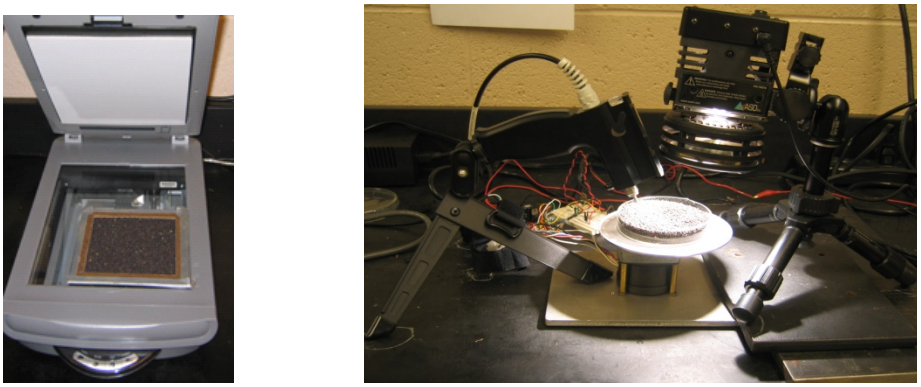
Figure 2.1. Scanned canola sample images with 0.1 (a) , 0.3 (b) , 1(c), 1.34 (d) , 1.45(e), 2.5 (f) , 3.4 (g), 4.3 (h), 4.6 (i) and 11.3% (j) total dockage

Spectral Data Acquisition

Spectral data, within the spectral range of 350nm to 2500nm, were collected using a FieldSpec Pro spectrometer (ASD Inc, CO) at a resolution of 1nm. Each sample was placed in a petri-dish on a turn-table. The diameter of the turn-table was 12.6 cm and rotated at 6 rpm. The speed of the turn-table could be varied using a potentiometer. A gear down stepper motor was used to control the speed of the turn-table.

Spectral measurements were taken using a fiber optic probe (ASD Inc, CO), with a 25 degree field of view, and an external light source, Pro Lamp (ASD Inc, CO). The probe was kept at a distance of 2.92 cm from the sample (Figure 2.2 b). The probe was recalibrated using a white standard having 99 percent reflectance (Spectralon Target, Labsphere, NH) before recording data for each sample. Since the probe was kept at an angle rotating the sample using a turn table blended the field of view of the spectrometer and eliminated scattering effects. The spectral data was used for validation of results

obtained from the scanner data, the data within the spectral range of 350nm to 800nm were used for further analysis.



(a) **(b)**

Figure 2.2. Set up for scanning canola samples (a) and for spectral data acquisition (b)

Data corresponding to wavelengths from 350-450 nm was not used in analysis due to the noise observed at the tails of the spectra. The samples used to collect the spectral data were the same as the ones used for scanner data. For each subsample, ten reflectance measurements were taken at an interval of 1 second. These measurements were then averaged to give the reflectance for each subsample. The reflectance measurements of 20 subsamples were then averaged to give the mean reflectance measurement for each sample.

Statistical analysis of both scanner and spectral data were performed using the SAS (SAS Institute Inc., version 9.2; NC, USA) statistical software. To grade canola, on the basis percent dockage present in the sample, the L, R, G and B domain data was standardized with the mean equal to zero and variance equal to 1 using PROC STANDARD procedure in SAS. The one way ANOVA (alpha =0.05) showed that atleast one sample was

significantly different in all the domains. The means of each sample were then compared by Tukey-Kramer HSD (Honestly Significant Difference) method using PROC GLM procedure at an alpha level of 0.05. According to this method, difference between means greater than or equal to the minimum significant difference are considered significantly different and the differences less than the minimum significant difference are considered to be insignificant.

Results and Discussion

The results obtained after comparing the means of each sample for L, R, G and B domains are shown in Table 2.1. Different lower case letters (shown after the data) show that the means of the corresponding samples are significantly different and same letters show that the means are not significantly different within the column in the table.

Sample 10 with 11.2% total dockage was significantly different from all other samples in all domains. According to Tukey's HSD method the samples have been categorized in eight groups for L and R domains, in seven groups for G domain and in five groups for B domain. Since sample 10 has comparatively higher percentage of total dockage present, there is a possibility that the results are affected by the presence of this sample. Therefore, the samples were re-analyzed after removing sample 10 from the analysis (Table 2.2).

Table 2.1 Mean comparison for each sample type using Tukey-Kramer HSD method for L, R, G and B domains, Lower case letters indicate significant differences within data column

Sample Number	Total Dockage	Mean Intensity Values			
		Luminosity	Red	Green	Blue
1	0.10%	40.46 (h)	42.07 (h)	40.07 (g)	38.01 (e)
2	0.30%	43.88 (e, f)	48.52 (d, e)	42.45 (d, e, f)	39.08 (d, e)
3	1%	41.35 (g, h)	43.7 (g, h)	40.79 (f, g)	38.02 (e)
4	1.34%	42.75 (f, g)	45.31 (f, g)	42.15 (e, f)	39.08 (d, e)
5	1.45%	45.72 (c, d, e)	50.53 (c, d)	44.29 (c, d)	40.45 (c)
6	2.50%	44.58 (d, e ,f)	47.25 (e, f)	44.02 (c, d, e)	40.35 (c, d)
7	3.40%	47 (b, c)	51.52 (c)	45.82 (b, c)	41.41 (b, c)
8	4.30%	48.58 (b)	53.77 (b)	47.2 (b)	42.12 (b)
9	4.60%	46.44 (c, d)	49.59 (c, d)	45.84 (b, c)	41.23 (b, c)
10	11.20%	57.96 (a)	64.69 (a)	56.56 (a)	47.54 (a)

Removing sample 10 improved the results especially in R domain. It was expected that the R domain would show maximum distinction in the samples because canola is reddish brown in color which is different from the color of the impurities present. In this domain all the samples were significantly different except for sample 5 (1.45% total dockage), which was grouped with samples 7 and 9. Samples in domains L, G and B were grouped in six, five and six groups respectively.

Further, linear discriminant analysis (LDA) was done in SAS using PROC DISCRIM procedure. In this method a discriminant /classification criterion is determined by a measure of generalized square distance. The criterion takes into account either with-in

group covariance (quadratic function) or the pooled covariance matrix (linear function). Prior probabilities of the groups are also considered (Multivariate Analysis: Discriminant Analysis, SAS 2009).

Table 2.2 Mean comparison for each sample type (excluding sample10) using Tukey-Kramer HSD method for L, R, G and B domains. Lower case letters indicate significant differences within data column.

Sample Number	Total Dockage	Mean Intensity Values			
		Luminosity	Red	Green	Blue
1	0.10%	40.46 (f)	42.07 (h)	40.07 (e)	38.01 (f)
2	0.30%	43.88 (d)	48.52 (d)	42.45 (d)	39.08 (e)
3	1%	41.35 (f)	43.7 (g)	40.79 (e)	38.02 (f)
4	1.34%	42.75 (e)	45.31 (f)	42.15 (d)	39.08 (e)
5	1.45%	45.72 (c)	50.53 (b, c)	44.29 (c)	40.45 (c, d)
6	2.50%	44.58 (d)	47.25 (e)	44.02 (c)	40.35 (d)
7	3.40%	47 (b)	51.52 (b)	45.82 (b)	41.41 (a, b)
8	4.30%	48.58 (a)	53.77 (a)	47.2 (a)	42.12 (a)
9	4.60%	46.44 (b, c)	49.59 (c)	45.84 (b)	41.23 (b, c)

The results obtained after applying LDA are shown with the help of the classification table, Table 2.3. The samples which are categorized as other samples are considered as misclassified samples. It is observed that only samples 1 and 10 are accurately classified. Sample 6 with 2.5% total dockage was highly misclassified. It showed a misclassification rate of 50%. It was mainly misclassified into samples 3, 4 and 9 with 1, 1.34 and 4.6% total dockage. Total error rate of the classification of the samples is approximately 14%.

A possible reason for the misclassification could be that when a sample comprising canola seed and foreign material was placed on the scanner there is a possibility that the impurities were not

captured in the scanned image. It was expected that the R domain would show significant difference between the samples as the color of the impurities such as grass or straw is significantly different from canola, which is reddish brown in color. But the samples had a mixture of impurities (straw, pieces of wood, dead grass, ergot, other grains like wheat, stones, insects etc) commonly found in canola samples instead of only grass or straw. Therefore, there is a possibility that the type and quantity of impurity present in the sample, when it was scanned, could have an effect on the observed results. As an example, if a sample had more of ergot/stones than grass/straw, it would resemble more like a sample with lesser amount of total dockage because the color of ergot or small stones might not be easily distinguishable from the color of canola. Also since the scanned sample was determined by weight therefore, weight of each kind of impurity would have an effect on the amount of visible impurities in scanned image.

Table 2.3 Classification Table for different canola samples* obtained after applying Linear Discriminant Analysis

Sample #	Total Dockage	0.10 %	0.30 %	1 %	1.34 %	1.45 %	2.50 %	3.40 %	4.30 %	4.60 %	11.20 %
1	0.10%	20	0	0	0	0	0	0	0	0	0
2	0.30%	0	16	0	0	4	0	0	0	0	0
3	1%	0	0	17	3	0	0	0	0	0	0
4	1.34%	0	0	1	17	0	2	0	0	0	0
5	1.45%	0	3	0	0	15	0	0	2	0	0
6	2.50%	1	0	4	2	0	10	0	0	3	0
7	3.40%	0	0	0	0	0	0	19	0	0	1
8	4.30%	0	0	0	0	1	0	0	19	0	0
9	4.60%	0	0	0	0	0	1	0	0	19	0
10	11.20%	0	0	0	0	0	0	0	0	0	20

* Number of samples for each type = 20

To overcome the possible error caused due to presence of multiple kinds of impurities, a set of experiments were designed where only one kind of impurity was introduced in the canola samples. The amount of impurities was introduced on the basis of weight therefore grass/straw because of their low weight and high volume were not used. To simulate the presence of straw and other impurities, wheat seeds were used instead. Wheat seeds being light in color are easily distinguishable and being heavier than grass would have lesser volume. Samples were prepared with 0, 1, 1.5, 2, 2.5 and 5% wheat seed impurity. Image and spectral data were recorded using the same procedure as described in the material and methods section. A linear trend was observed for the mean L, R, G and B values obtained from scanner data with the R-square value of around 0.98 for L, R and B domains. G domain showed R-square value of 0.87 (Figure 2.3).

The Tukey's HSD method grouped the samples broadly into 4 groups in L, R and B domains. For G domain the samples were categorized mainly in three groups. Misclassification rate of 60% was observed when data was analyzed through discriminant analysis. Therefore, it can be concluded that apart from the different properties of different impurities, the way the samples are scanned also has an effect on the final classification results. Canola seeds being smaller in size tend to move below the impurities (wheat seeds) when the sample is spread on the scanner. As a result the image captured would show lesser amount of total dockage then actually present.

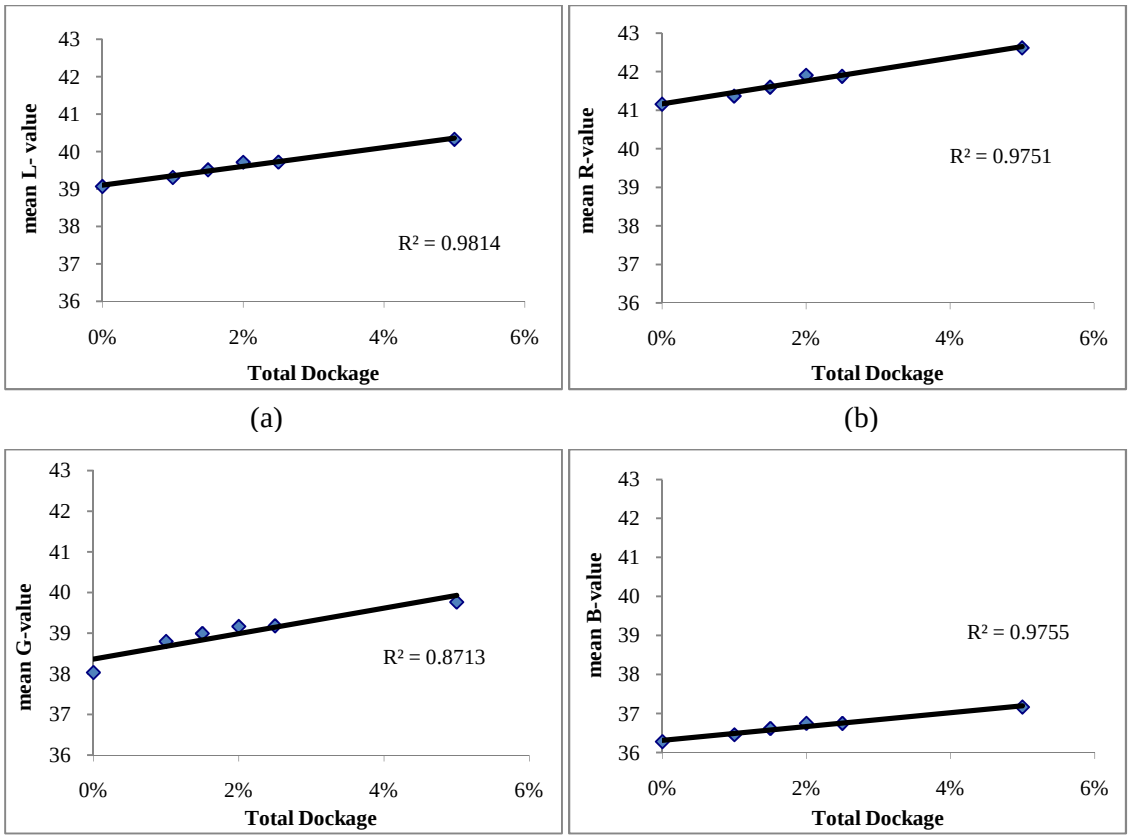


Figure 2.3 Mean L (a), R (b) , G (c) and B (d) values when wheat seeds are added as impurities in canola samples

Spectral data were used to compare the results obtained from scanner data as both these techniques are based on the standard reflectance of an object. Though, there appears to be difference in the spectral pattern with change in foreign material in the sample, no definite trend can be observed. It was expected that the scanner data would show some direct correlation with percent reflectance of different samples. Figure 2.4 shows the percent reflectance measurement of each sample. Difference in spectral measurements with change in total dockage can be observed in the red region (around 650 nm). It is very difficult to distinguish between samples with less than 4.6% total dockage in the blue (475 nm) and green (510 nm) regions. Mean spectral data for blue (475 nm), green

(510 nm) and red (650 nm) wavebands were compared using Tukeys HSD method (Table 2.4).

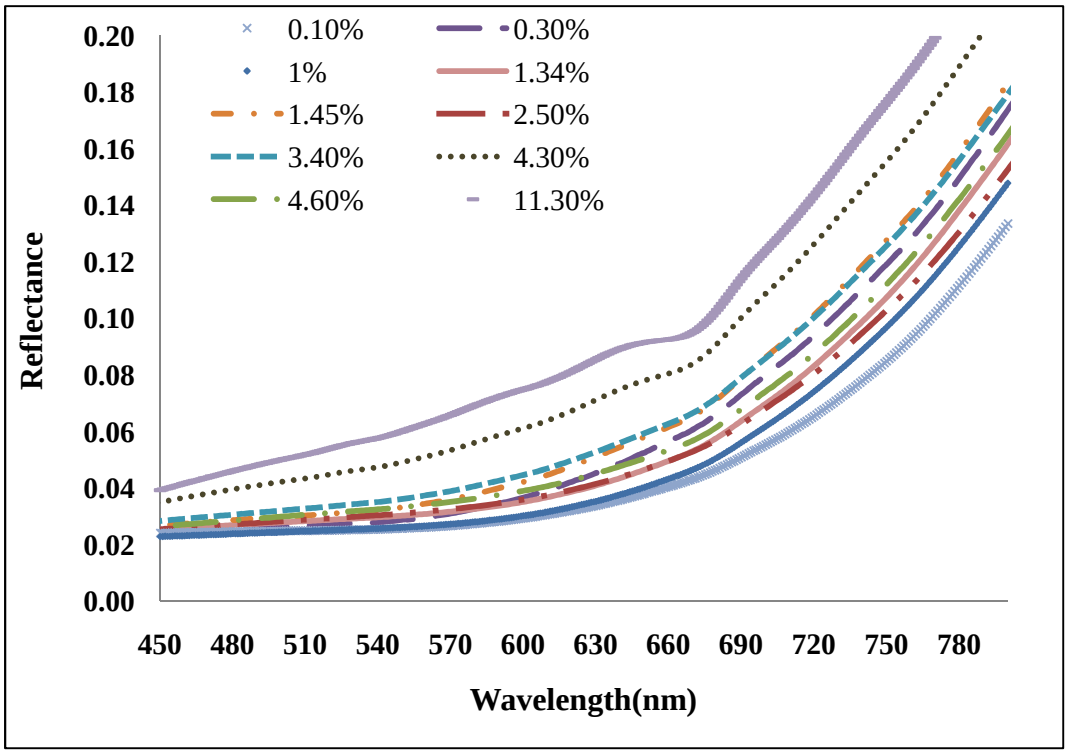


Figure 2.4 Mean reflectance data collected for 10 samples of canola differentiated by the amount of impurities in each

Sample 10 was significantly different from all other samples. This was expected, since sample 10 has maximum amount of total dockage in it. Sample 8 is also significantly different from all other samples in green and red regions. Sample 9 was significantly different in blue region. All other samples are similar to one or the other sample. Similar trends were also observed in scanner data.

Table 2.4 Mean comparison for each sample type for Blue (475 nm), Green (510 nm) and Red (650 nm) wavebands. Lower case letters indicate significant differences within data column

Sample Number	Total Dockage	Mean Intensity Values		
		475	510	650
1	0.10%	0.024 (e)	0.025 (e)	0.038 (f)
2	0.30%	0.026 (e, d)	0.027 (e, d)	0.053 (c, d)
3	1%	0.024(e)	0.025 (e)	0.04 (e, f)
4	1.34%	0.027 (c, d, e)	0.028 (c, d, e)	0.046 (d, e)
5	1.45%	0.028 (c, d)	0.03 (c, d)	0.058 (c)
6	2.50%	0.027 (c, d, e)	0.029 (c, d, e)	0.047 (e, d)
7	3.40%	0.03 (c)	0.033 (c)	0.059 (c)
8	4.30%	0.039 (c, d)	0.043 (b)	0.078 (b)
9	4.60%	0.028 (b)	0.031 (c, d)	0.051 (d)
10	11.20%	0.045 (a)	0.052 (a)	0.091 (a)

It can be concluded that the samples cannot be classified, on the basis of total dockage present, by using color features alone. Apart from variation caused by different impurities, variation in the color of canola seed would also have an effect on the reflected values. Visible range of the spectrum only takes into account the physical characteristics of the sample. The physical parameter such as color may get affected by the concentration of chlorophyll and other similar compounds. This causes variation in seed color, thus affecting results.

Therefore, only surface color does not provide sufficient information for the classification of the samples. Similar results were also reported by Paliwal et al (2003). It was concluded that a combination of morphological, textural and color features is required to obtain better accuracies of the classification model.

Conclusions

Though extensive research has been conducted to size, classify and grade rice, wheat, pulses, soybeans and lentils using machine vision, little work has been done for grading canola. The objective of this study was to grade canola samples on the basis of total dockage. Different samples with varying degree of total dockage were prepared and tested. Mean values of luminosity (L), red (R), green (G) and blue (B) were recorded for 20 sub-samples of each sample type.

The means were compared using Tukey-Kramer HSD method. According to the results obtained after mean comparison, only sample 10 with 11.2% total dockage was significantly different from all the other samples. The discriminant analysis showed a misclassification rate of 14%. The sample with 2.5% total dockage showed maximum error during classification of samples. This method showed some potential to discriminate between samples on the basis of amount of total dockage present, but was not able to give conclusive results. The results obtained from spectral data were similar to the scanner data and confirmed the findings.

References

Bal Satish, T.K.Basu, Maqsood Ali and P.C. Proddutur. 2006. The determination of morphological and textural features of rice grains in a sample using digital image processing technology and the classification of the rice grains in the sample. Presented at ASABE Annual International Meeting, Portland, Oregon, USA. Paper number: 066009.

Berglund D.R., K. McKay, and J. Knodel. 2007. Canola Production: A-686, August 2007. Available at: http://www.ag.ndsu.edu/pubs/plantsci/crops/a686w.htm. Accessed on: 2 February 2011.

Boyles Mark. Canola Project Specialist, OKANOLA, Oklahoma State University. Available at: http://www.canola.okstate.edu/commonquestions/index.htm. Accessed on: 2 Feburary 2011.

Casady W.W., M.R. Paulsen, J.F. Reid, and J.B. Sinclair. 1992. A trainable algorithm for inspection of soybean seed quality. Transactions of ASABE. 35(6):2027-2034.

Delwiche S.R. 2003. Classification of Scab- and other Mold- Damaged Wheat Kernels by Near-Infrared Reflectance Spectroscopy. Transactions of the ASAE. 46(3):731-738.

Kumar P.A. and S. Bal. 2006. Non destructive quality control of rice by image analysis. Presented at ASABE Annual International Meeting, Portland, Oregon. Paper number: 066083.

Luo, X., D. S. Jayas, and S. J. Symons. 1999. Identification of damaged kernels in wheat using a colour machine vision system. Journal of Cereal Science. 30(1):49-59.

Majumdar S., and D. S. Jayas. 1999. Classification of bulk samples of cereal grains using machine vision. Journal of Agricultural Engineering Research. 73(1):35-47.

Majumdar S., and D. S. Jayas. 2000. Classification of bulk samples of cereal grains using machine vision: I. Morphology Models. Transactions of ASAE. 43(6):1669-1675.

Nair Manikandan.1997. Dockage identification in wheat using machine vision. MS thesis. Winnipeg, Canada: University of Manitoba, Department of Biosystems Engineering.

NG H. F., W. F. Wilcke, R. V. Morey, and J.P. Lang. 1998. Machine vision evaluation of corn Kernel mechanical and mold damage. Transactions of the ASABE. 41(2):415-420.

Paliwal J., N.S. Shashidhar, D.S Jayas. 1999. Grain kernel identification using kernel signature. Transactions of ASAE. 42(6):1921-1924.

Paliwal, J., N. S. Visen, D. S. Jayas, and N. D. G. White. 2003. Cereal Grain and Dockage Identification using Machine Vision. Biosystems Engineering. 85 (1):51-57.

Paliwal J., D.S. Jayas, N.S. Visen and N.D.G. White. 2004a. Feasibility of a machine vision based grain cleaner. Applied Engineering in Agriculture ASAE. 20(2):245-248.

Paliwal J., M.S. Borhan and D.S. Jayas. 2004b. Classification of cereal grains using a flatbed scanner. Canadian Biosystems Engineering. 46:3.1-3.5.

SAS. 2009. SAS User's Guide: Statistics. Ver. 9.2. Cary, N.C.: SAS Institute, Inc.

Shahin Muhammad A. and S.J. Symons. 2001. Lentil seed size distribution with machine vision. Presented at ASAE Annual International Meeting, Sacramento, California, USA. Paper number: 013058.

Shahin Muhammad A., S.J. Symons and A. X. Meng. 2004. Seed sizing with image analysis. Presented at ASAE/CSAE Annual International Meeting, Ottawa, Ontario, Canada. Paper number: 043121.

Shahin Muhammad A., S.J. Symons and V.W. Poysa. 2006. Determining soyabean seed size uniformity with image analysis. Biosystems Engineering. 94(2):191-198.

Shatadal P. and J. Tan. 2003.Idenifying damaged soybeans by color image analysis. Applied Engineering in Agriculture ASAE. 19(1):65-69.

Tahir A.R., s. Neethirajan, D.S. Jayas and J. Paliwal. 2006. Assessment of machine vision algorithm for quantification of foreign matter in wheat. Presented at ASABE Annual International Meeting, Portland, Oregon, USA. Paper number: 063087.

Tanska Malgorzata, Daniela Rotkiewicz, Witold Kozirok, and Iwona Konopka. 2005. Measurement of the geometrical features and surface color of rapeseeds using digital image analysis. Food Research International. 38 (7):741-750.

Visen N.S., J. Paliwal, D. Jayas, N.G. White. 2003. Image analysis of bulk grain samples using neural networks. Presented at ASAE Annual International Meeting, Las Vegas, Nevada, USA. Paper number: 033055.

Wang D., M.S. Ram and F.E. Dowell. 2002. Classification of Damaged Soybean Seeds Using Near-Infrared Spectroscopy. Transactions of the ASAE. 45(6):1943-1948.

CHAPTER III

IDENTIFICATION OF DISTINCTLY GREEN AND HEAT TREATED SEEDS USING MACHINE VISION

Abstract: According to USDA standard guidelines, US grade 1 canola should not have more than 3% of total damaged seeds (distinctly green seeds and heat treated seeds). Distinctly green seeds should not be more than 2% and heat treated seeds should not be more than 0.1%. Graders usually identify these seeds by crushing a sample of canola and inspecting the crushed seeds visually. But visual inspection can introduce human error. Therefore, there is a need to develop an automated system that can distinguish these seeds from the desired canola seeds. Machine vision can be a potential method to achieve this objective as similar technology has also been applied on wheat and soybean to classify damaged seeds. The results from this study indicated 100 percent classification accuracy for distinctly green seeds but some of the good and heat damaged seeds were misclassified. It was concluded that machine vision can be a potential method for grading of canola on the basis of seed color.

Review of Literature

Machine vision, image spectroscopy and near infrared (NIR) spectroscopy are among a few of the techniques being used to grade different grain varieties. Machine vision involving identification and separation of grains by digital image processing is one of the most common techniques. Impurities are separated from the grain using different physical

features like size, shape, texture, color and other morphological parameters. Morphological and textural parameters have been used to identify the quality of rice (Bal *et al*, 2006). High-resolution images, obtained using three chip charge coupled device (CCD) color cameras, have been successfully used to identify different grain kernels. This method uses an algorithm based on kernel signature which involves shape, length and color of different grain kernels. Though this algorithm was able to identify different grain kernels but classification of damaged kernels and foreign material content was not tested (Paliwal *et al*, 1999). Image analysis in conjunction with a back propagation neural network as a classifier has been used to identify different grain types on the basis of color and textural features (Visen et al, 2003). Another study on quantification of foreign material (barley) in wheat has shown that back propagation neural networks with statistical classifiers can be used to classify wheat and barley admixture correctly. The neural network showed better performance than the statistical classifier. A need to improve the algorithm to improve the efficiency of the classifier was identified because the classification accuracy was less for barley admixture equal to 1.2% (Tahir et al, 2006). Machine vision has also been used to quantify the percentage of dockage material in the grain sample before and after it has passed through the cleaner and thus has been used to test the performance of a grain cleaner (Paliwal, 2004). The discussed studies have used the CCD camera as their image acquisition device.

CCD cameras yield high resolution images but are quite expensive. A rather inexpensive machine vision system that has been used and tested by researchers is the flatbed scanner (Paliwal et al, 2004; Shahin et al, 2001; Shahin et al, 2006). Flatbed scanners with back propagation neural networks as a classifier have been successfully used to classify cereal

grains using color, textural and morphological features of the samples (Paliwal et al, 2004). Machine vision techniques using flat bed scanners have also been applied for determining the seed size distribution of lentil seeds, seed sizing of pulses, color and size grading of pulse grains, seed size uniformity of soybean seeds, and quality of rice (Shahin et al, 2001; Shahin et al, 2004; Shahin et al, 2006; Kumar and Bal, 2006). Researchers have used thresholding based on histogram analysis in conjunction with other morphological features to segment damaged seeds (Shatadal and Tan, 2003) in soybeans.

United States grain grading standards indicate that the presence of distinctly green (DG) and heat damaged (HD) seeds should not be more than 3, 10 and 20 percent in US grade 1, 2 and 3 canola respectively. The traditional method of identifying these seeds is through visual inspection. A strip containing 500 seeds is prepared using canola counting paddle with 500 depressions (Figure 3.1). The seeds are then crushed using a small wall paper roller and the color of crushed seeds is then compared with the grading mat for canola, provided by United States Department of Agriculture (USDA). However, this method is subjective in nature and human error is possible. Identifying these seeds using machine vision is the main objective of this study.

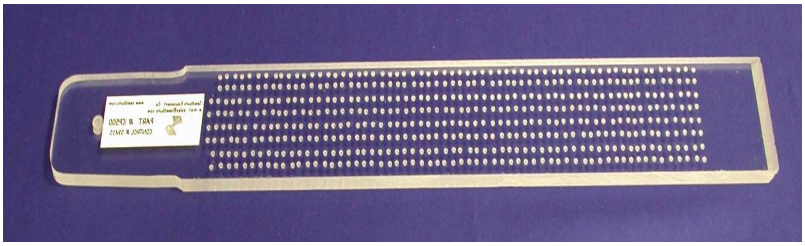


Figure 3.1. Canola counting paddle (picture taken from seedburo.com)

Material and Methods

Sample Preparation

Using a 10 gm sample of canola, the depressions of canola counting paddle were filled and a seed strip as described in previous section, was prepared. For initial testing, some of the seeds were removed and replaced by heat-treated and distinctly green seeds. The Enid Grain Inspection Company, Inc., Enid, Oklahoma, provided the HD and DG (damaged kernels) seed samples. The seeds were then crushed using a small wallpaper roller and the strip was scanned using a flatbed scanner (Epson Perfection V500 photo scanner, Epson America Inc., Long Beach, CA, USA). Seeds were removed from the alternate rows of the counting paddle to avoid joining of seeds after crushing (Figure 3.2). The MATLAB Digital Image Processing toolbox (Mathworks, Natick, MA, USA) was used to analyze the scanned images. Image segmentation in the B domain was done to distinguish seeds from the background based on color. The RGB image of the seed was converted to L*a*b domain and data for both domains was recorded. The seeds were segmented out from the background and labelled using morphological operations like successive dilation and erosion. Different properties such as pixel value, area and centroid for each labelled seed were calculated for enhanced and non-enhanced images for both the domains. Image contrast enhancement helped in providing better classification criterion of seeds based on their color.

Figure 3.2. Seed strip prepared for scanning and grading

The seeds with very small areas were removed. This was done by calculating outlier area values using mean and standard deviation. Seeds having areas that were very small as compared to other seed areas were deleted. The seeds were re-labelled and their mean RGB and L*a*b values were calculated and recorded. The 'L' value in the L*a*b color space indicates level of lightness or darkness. It ranges from 0 to 100 with 100 indicating light and 0 indicating dark color. The value of 'a' indicates the level of redness or greenness. A positive value of 'a' is an indicator of redness and its negative value indicates greenness. The positive 'b' value indicates yellowness and its negative value indicates blueness (HunterLab, 2008). Presence of the brown coating of seeds (brown pixels) mixed with crushed seed created bias in the values obtained from the scanned segmented image. Therefore, another thresholding operation was applied in the red domain of each seed to correct this bias. Thresholding was done in the red domain because the maximum difference between brown and the remainder of the pixels was observed in this domain. The pixels with less than the threshold value (mean value of brown pixels) were deleted and the mean of the remaining pixels were calculated for both

the color spaces. The mean values of each seed were compared to develop a classification criterion to classify seeds into good, DG and HD seeds. The algorithm developed is given in Appendix I.

Results and Discussion

Since no standard values in the respective color spaces are available for DG, HD and good seeds therefore, data for individual DG, HD and good seed was compared to decide the grading criterion. For this 5 seed strips similar to one shown in Figure 3.2 were prepared. Out of these, data from the 2 strips were used to develop the classification criterion for good and HD seeds. A total of 84 good seeds and 47 HD seeds were used from these strips. When the introduced DG seeds were crushed using the wall paper roller it was observed that most of the seeds had matured and were not distinctly green in color. Some marginal cases were also observed in the DG seeds. The seeds were compared visually with the standard charts provided by the USDA and categorized as DG or good seeds. Total number of DG seeds were very few in quantity, thus data from 4 strips was used to develop its classification criterion. Total number of DG seeds used for developing the algorithm was four. Since DG and HD seeds were introduced after replacing the good seeds on the paddle therefore, number of DG, HD and good seeds were known in each strip. The classification criterion was based on the maximum and minimum values of RGB and L*a*b values of both original and contrast enhanced images (Tables 3.1 and 3.2).

It was observed that the ‘a’ value for DG seeds was negative and for HD seeds positive. The negative ‘a’ value indicates the degree of greenness of DG seeds. Good seeds covered both the negative and the positive range of the ‘a’ value but their mean value was negative for the original image. The variation in value of ‘a’ for good seeds can be attributed to the maturity of the seed. For the contrast enhanced image ‘a’ value for DG seeds was very small and ranged from 0.0048 to 0.067 with mean value equal to 0.022, whereas for good seeds this value ranged from 0.02 to 0.6 with mean value equal to 0.26. For HD seeds the ‘a’ value for contrast enhanced imaged ranged between 0.23 and 0.84 with mean value equal to 0.61. The ‘a’ value in both image types appeared to show significant differences.

Table 3.1. Intensity values of different seed type in RGB and L*a*b color spaces for original image

Seed Type		R	G	B	L	a	b
	mean	182.6	105.4	159.3	83.2	-1.0	19.3
	min	164.9	83.7	139.8	79.4	-4.4	11.8
Good seed	max	195.3	123.3	173.3	85.7	1.8	28.4
	mean	174.1	99.4	143.9	80.3	1.5	17.7
	min	164.1	89.6	133.8	78.1	-0.1	12.5
HD* seed	max	185.8	116.5	159.2	83.4	2.8	21.2
	mean	170.0	87.9	156.3	81.9	-4.6	24.9
	min	163.7	82.2	153.5	81.3	-5.6	22.2
DG** seed	max	176.3	93.6	162.3	83.1	-3.7	27.6

*HD: Heat Damaged; ** DG: Distinctly Green

The means for both the color spaces and images were compared using Tukey-Kramer highest significant difference (HSD) method in JMP statistical software (version 7.0.2, SAS Institute Inc., Cary, NC, USA). Significant differences between the samples were observed when the means were compared using R, G, L, ‘a’ and ‘b’ data for original images. For contrast enhanced images data in ‘B’ and ‘a’ domains showed significant

differences. Based on the results obtained after comparing the means it was concluded that the value of 'a' had a significant effect on the classification of seeds. Therefore classification criterion was developed based on the RGB and 'a' values. Results obtained after testing all the seed strips using the developed classification criterion are shown in Table 3.3.

Table 3.2. Intensity values of different seed type in RGB and L*a*b color spaces for contrast enhanced image

Seed Type		R	G	B	L	a	b
Good seed	mean	161.1	92.4	144.3	1.0	0.1637	1.0
	min	140.9	57.5	124.9	1.0	0.0000	1.0
	max	182.8	122.2	162.5	1.0	0.6287	1.0
HD* seed	mean	150.5	91.0	127.6	1.0	0.6127	1.0
	min	136.3	74.9	118.0	1.0	0.2268	1.0
	max	162.5	134.3	143.1	1.0	0.8398	1.0
DG** seed	mean	149.3	84.4	151.4	1.0	0.0205	1.0
	min	140.8	66.1	143.5	1.0	0.0048	1.0
	max	158.1	97.2	159.1	1.0	0.0674	1.0

*HD: Heat Damaged; ** DG: Distinctly Green

Table 3.3. Classification results obtained after machine vision analysis

Seed Type	Strip 1		Strip 2		Strip 3		Strip 4		Strip 5	
	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted	Actual	Predicted
DG** Seed	0	0	1	1	1	1	3	3	1	1
HD* Seed	20	22	19	22	28	22	12	11	12	14
Good Seed	96	94	96	93	87	93	60	61	74	72

*HD: Heat Damaged; ** DG: Distinctly Green

Classification accuracy of 100% was observed for DG seeds. Some of the HD seeds were classified into good seeds and some of the good seeds were misclassified into HD seeds. Classification accuracy of HD seeds ranged from 80-90% (approximately). The misclassification was mainly because of the intensity values for both HD and good seeds were overlapping each other. A possible reason for this could be that the base color for

both these seed types is similar, HD seeds are darker than good seeds. Inaddition, there is a possibility that some seeds are marginal and cannot be categorized into HD or good seeds by visual inspection. Also presence of brown pixels due to upper coating of crushed canola seeds would affect the overall intensity values of the seeds. Though care had been taken to remove these pixels before calculating the average intensity values of the seed but presence of few of these pixels cannot be ruled out thus overall classification efficiency of the algorithm is affected.

Conclusion

The results obtained from this study indicated that machine vision can be used to distinguish between good and damaged kernels. The distinctly green seeds were identified with 100 percent accuracy but some inaccuracies were observed in the classification of good and heat damaged seeds. Few of the heat damaged seeds were classified as good seeds and vice versa. This study indicates the possibility of use of machine vision as a grading method but more testing of the algorithm is suggested.

References

Bal Satish, T.K.Basu, Maqsood Ali and P.C. Proddutur. 2006. The determination of morphological and textural features of rice grains in a sample using digital image processing technology and the classification of the rice grains in the sample. Presented at ASABE Annual International Meeting, Portland, Oregon, USA. Paper number: 066009.

Grain Inspection handbook. USDA Grain Inspection, Packers and Stockyards Administration, Federal Grain Inspection Service.

HunterLab Application Note .2008. Hunter L,a,b versus CIE 1976 L*a*b. Insight on color . 13 (2).www.hunterlab.com/appnotes/an02_01.pdf. Accessed on 23 September 2010.

Kumar P.A. and S. Bal. 2006. Non destructive quality control of rice by image analysis. Presented at ASABE Annual International Meeting, Portland, Oregon. Paper number: 066083.

Paliwal J., N.S. Shashidhar, D.S Jayas. 1999. Grain kernel identification using kernel signature. Transactions of ASAE. 42(6):1921-1924.

Paliwal J., D.S. Jayas, N.S. Visen and N.D.G. White. 2004. Feasibility of a machine vision based grain cleaner. Applied Engineering in Agriculture ASAE. 20(2):245-248.

Paliwal J., M.S. Borhan and D.S. Jayas. 2004. Classification of cereal grains using a flatbed scanner. Canadian Biosystems Engineering. 46:3.1-3.5.

Shahin Muhammad A. and S.J. Symons. 2001. Lentil seed size distribution with machine vision. Presented at ASAE Annual International Meeting, Sacramento, California, USA. Paper number: 013058.

Shahin Muhammad A., S.J. Symons and A. X. Meng. 2004. Seed sizing with image analysis. Presented at ASAE/CSAE Annual International Meeting, Ottawa, Ontario, Canada. Paper number: 043121.

Shahin Muhammad A., S.J. Symons and V.W. Poysa. 2006.Determining soya bean seed size uniformity with image analysis. Biosystems Engineering. 94(2):191-198.

Shatadal P. and J. Tan. 2003.Idenifying damaged soybeans by color image analysis. Applied Engineering in Agriculture ASAE. 19(1):65-69.

Tahir A.R., s. Neethirajan, D.S. Jayas and J. Paliwal. 2006.Assessment of machine vision algorithm for quantification of foreign matter in wheat. Presented at ASABE Annual International Meeting, Portland, Oregon, USA. Paper number: 063087.

Visen N.S., J. Paliwal, D. Jayas, N.G. White. 2003. Image analysis of bulk grain samples using neural networks. Presented at ASAE Annual International Meeting, Las Vegas, Nevada, USA. Paper number: 033055.

CHAPTER IV

NON DESTRUCTIVE CLASSIFICATION OF CANOLA ON THE BASIS OF ERUCIC ACID USING NIR SPECTROSCOPY

Abstract: Canola seeds look similar to mustard and rapeseed. Both mustard and rapeseed have erucic acid around 40- 60% of total fatty acids, which makes them unhealthy especially for people suffering from high cholesterol. Inaddition, these seeds act as impurities in canola. According to the US standard guidelines canola seed should have less than 2% erucic acid. Conventionally amount of erucic acid has been determined using gas chromatography, which is an expensive and a time consuming procedure. Researchers have been successful in determining various free fatty acids, like linoleic, linolenic and oleic acids, using NIR spectroscopy. NIR spectroscopy has also shown potential in estimating erucic acid. This study focuses on classification of canola seed from other high erucic acid seeds on the basis of erucic acid using NIR spectroscopy. Classification of seeds was done using linear discrimnant analysis on the principal components obtained after applying principal component analysis on the spectral data. Seeds were classified broadly into six classes (less than 1% to 52.6% erucic acid) with 100% accuracy. Sensitivity of the analysis was also tested by analyzing canola samples with different levels of erucic acid (0 to 1.02% erucic acid). A100% accuracy was achieved when a test dataset was tested on the calibration model. Wavebands at 1690, 1720, 1756, 1769 and 1860 nm showed good response to change in erucic acid levels and thus, can be considered as potential wavelengths that can be used to estimate erucic acid.

Review of Literature

US standard guidelines indicate that canola should have less than 2% erucic acid. Most of the other seeds in *Brassica* family, like mustard and rapeseed have erucic acid of around 40-60% and are considered as impurities in canola. It is difficult to visually distinguish these seeds from canola seeds as these look similar, especially rapeseed. Past research suggests the use of near-infrared radiation (NIR) spectroscopy in estimating erucic acid in rapeseed and mustard. Near-infrared radiation covers the region from 700 – 2500 nm in the electromagnetic spectrum. It mainly responds to vibration energy levels of a molecule. The type of vibrations of a particular functional group determines absorption spectrum in the infrared region of the electromagnetic spectrum. Functional groups O-H, C-H, C-OOH, C=O, aromatic C-H groups and N-H are mainly identified in the NIR region. Thus food quality, which is affected by amount of moisture content, organic acids, food oxidative products, sugars etc present in it, can be measured using NIR spectroscopy (Sundaram et al., 2009). NIR spectroscopy can be used as a non destructive technique for oilseed analysis (Daun and Williams, 1997). It has been successfully used to estimate the oil content in whole seed canola (Greenwood *et al,* 1999). The developed calibration model was based on the samples analysed from 1995 to 1997. Stability of the model still needed to be evaluated to predict oil content with future years of canola samples. Sato et al (1998) had successfully evaluated the feasibility of using NIR spectroscopy as a non-destructive method for estimation of fatty acid composition of rapeseed. Velasco and Becker (1998) were able to develop highly accurate calibrations for oleic, linoleic, linolenic and erucic acid for rapeseed using 3 g sample. In another study, Velasco, Martinez and Haro (1997) identified six spectral regions: 1140-1240,

1350-1400, 1650-1800, 1880-1920, 2140-2200 and 2240-2380 nm to estimate the fatty acid composition of oil in intact seed samples of Ethopian mustard. Velasco *et al*(1999) conducted a study to develop calibration equations to estimate seed oil content and concentration of major fatty acids for intact seeds samples of Brassicaceae family. It was concluded that only oil content and concentrations of $C_{18:1}$, $C_{18:3}$ and $C_{22:1}$ can be estimated reliably by using NIR spectroscopy. Calibrations for non-destructive estimation of erucic acid and glucosinate content in rapeseed and mustard seeds have been developed by using partial least square regression on data obtained from Fourier transform near infrared reflectance spectroscopy (FT-NIR). A close relationship was observed between reference method and predicted values from the spectral data from 7502.1 to 5444.6 cm^{-1} (Kumar et al., 2010).

NIR spectroscopy was also applied to determine the fatty acid (FA) composition of husked sunflower (Sato *et al* 1995). In one of the studies it was observed that the accuracy of the calibration model based on NIR spectroscopy to estimate oil content and fatty acid decreased when intact (non husked) sunflower seed was used as compared to the accuracy obtained from oil, meal or husked seed (Pérez-Vich et al., 1998). Seimens and Daun (2005) used partial least square (PLS) analysis to develop calibration model for determining fatty acid composition of canola, flax and solin. It was concluded that NIR spectroscopy showed potential in estimating linolenic acid, oleic acid and total saturated FA. Pallot *et al* (1999) also developed calibrations for eight major fatty acids (C16:0, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1, C22:1) using full NIR spectral information instead of selected wavelengths, for each separate group of samples (canola, high erucic acid rapeseed, canola quality and condiment mustard samples). Nuclear magnetic

resonance (NMR) spectroscopy has also been evaluated as a non-destructive technique to develop seed oil calibrations (Hutton, Garbow and Hayes; 1999).

Apart from non-destructive estimation of fatty acid composition NIR spectroscopy has also been used for non-destructive discrimination between rapeseed cultivars (Zou et al., 2011). This was done using distance discriminant analysis (DDA) and back propagation neural network (BPNN). DDA and BPNN were applied on the first six principal components (PCs) obtained after applying principal component analysis (PCA) on the spectral data. In this study 225 samples from 5 kinds of rapeseed were used. 100% prediction accuracy was achieved by both the methods but it was concluded that DDA was a better approach over BPNN since it was convenient and more intuitive. Another study showed an overall correct identification rate of 96.34% when transgenic rapeseed oil was distinguished from non transgenic rapeseed oil (Zhu et al., 2011). This study used PCA in conjunction with discriminant partial least squares (DPLS). PCA score plot showed that both oil types can be distinguished from each other roughly. A similar study on *Pinus sylvestris* L. based partial least squares discriminant analysis (PLS-DA) showed a 100% classification accuracy when seeds were identified on the basis of seed sources and parents using visible (VIS) and NIR spectroscopy (Tigabu et al., 2005). Discriminant analysis on NIR data has also been successfully used to classify four different vegetables oil types (cotton-seed, peanut, soybean and canola). Four wavelengths, 1704,1802,1816 and 2110 nm were identified and were used to build the model (Bewig et al., 1994).

Based on past research work, non destructive classification of canola on the basis erucic acid using NIR spectroscopy, was outlined as the main objective for this study. Sensitivity of the classification technique would also be evaluated.

Material and Methods

Samples

A total of 192 samples of seven types were analyzed. The variation in erucic acid ranged from less than 1% to 52.6 % of total fatty acids (Table 4.1). Erucic acid is expressed as percent total fatty acids per gram of oil. Sample 1 was a combination of 8 canola samples acquired from local farmers and stored for different periods of time. The range of erucic acid in sample 1 was from 0 to 1.05 % with a standard deviation (SD) of 0.38. Sample 1 (Table 4.2) was labeled as sample with less than equal to 1% erucic acid. Samples A, B and C (Table 4.2) were part of canola obtained from a local farmer in 2007. Sample 'A' was tested when it was received that is on Day 0. Sample 'B' was stored in an oven at 45 degree C for a period of 25 days so as to change it chemical properties. Sample 'C' was obtained by storing under natural hot and humid conditions in a grain bin for a period of 60 days. Samples 'D', 'E', 'F' and 'G' (Table 4.2) were obtained from a local farmer in 2009. Sample'D' was tested at day 0 when it was received. Sample 'E' was kept under natural hot and humid conditions for 30 days. Since weather conditions changed therefore rest of canola sample was shifted to an environment chamber where samples 'F' and 'G' were stored at 40 degree C for 7 and 8 months respectively before getting tested. Sample 'H' was new canola bought from a local vendor in 2010.

Samples 2, 3, 4 and 5 were mixtures of canola and rapeseed in different proportions (7:1, 3:1, 1:1 and 1:3 respectively, canola: rapeseed). Samples 6 and 7 were mustard (Angelina Gourmet, Canada) and rapeseed (Technology Crops International, Winston-Salem, NC)

respectively. Mixtures of canola and rapeseed were introduced to create variations in percent erucic acid with in the samples.

Table 4.1. Samples with different levels of erucic acid

Sample number	Erucic Acid Content (%)	Number of Samples
1	Less than equal to 1	140
2*	5.94	3
3*	10.75	3
4*	20.86	3
5*	33.48	3
6**	46.11	20
7***	52.6	20

* mixture of canola and rapeseed, ** rapeseed, *** mustard seed

Table 4.2. Sample 1 as a mixture of different canola samples

Sample 1	Erucic Acid Content (%)	Number of Samples
A	0.22	10
B	0.72	10
C	0.08	20
D	0.08	20
E	0.08	20
F	0.03	20
G	0	20
H	1.02	20

Spectral Data Acquisition

Spectral data, within the spectral range from 350nm to 2500nm was collected using a FieldSpec Pro spectrometer (ASD Inc, CO). It has a spectral resolution of 3nm at 700 nm with a sampling interval of 1.4 nm between 350 -1050 nm. The spectrometer has a spectral resolution of 10 nm at 1400 and 2100 nm with a sampling interval of 2 nm

between 1000- 2500 nm. The field of view was 25 degrees and reflectance was recorded from a height of 2.9 cm. Since the reflectance data was not taken at 90 degree angle the scattering of wavelengths was taken into account by rotating the sample using a turn table (Figure 4.1). The probe was recalibrated using a white standard having 99 percent reflectance (Spectralon Target, Labsphere, NH) before recording data for each sample. Due to high noise level at the ends of the spectra the regions from 300-400 nm and 2400-2500 nm were not used in the analysis. From past research it has been concluded that the NIR region of the electromagnetic spectrum responds to the changes in fatty acid composition. Therefore, data only in the NIR range of 1100 nm to 2400 nm was used for further analysis.

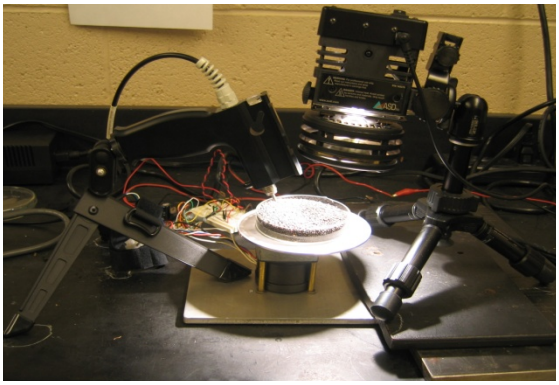


Figure 4.1. Set up for spectral data acquisition

Lab estimation of erucic acid

Tables 4.1 and 4.2 show the percent erucic acid, averaged for replications, present in each sample. Percent erucic acid is given as g of total fatty acid/ g of oil sample. 20 g samples

of each sample type were ground using a 2-L Waring blender at low speeds for 2 minutes. The blender was kept 'on' for 20 seconds and then was turned off. The seed was ground at 20 second interval for a total of six times such that total time the sample was ground was 2 minutes. The sample was ground for short intervals to avoid any heating of the sample. The sample was ground to pass through a 1 mm sieve.

Oil was extracted from the seed samples using ether extraction with a rotary evaporator. Three replications were done for each type of sample and extraction. The methyl esters of fatty acids, made by using methanolic HCL, were then used to determine fatty acid composition of each sample. Heptadecanoic acid (HDA) was used as an internal standard. Gas chromatography (GC) analysis was conducted using a Tracor model 540 gas chromatograph (Tracor Intsruments, Austin, TX) fitted with a DB 23 fused silica capillary column (30 m x 0.25 µm film thickness; J and W Scientific Inc., Rancho Cardova, CA) and a flame ionization detector. Helium was used as a carrier gas at a linear flow rate of 20 cm/sec. The fatty acid methyl esters were separated using a GC program: the initial temperature increased from 50 ^{O}C to 180 ^{O}C at 10 ^{O}C/min, a hold at 180 ^{O}C for 5 min, the temperature then increased linearly from 180 ^{O}C to 240 ^{O}C at 5 ^{O}C/min and a hold at 240 ^{O}C for a final 5 min period. Fatty acids compounds were identified and their peak areas were obtained using Spectra-Physics 4270 integrator (Spectra-Physics Inc., San Jose, CA). Peak area response was quantified relative to HDA.

Data Processing

Before analyzing the spectral data, the data was pre-processed using Unscrambler 9.0 (CAMO Software Inc, Woodbridge, NJ, USA). Pre-processing of data included transformation of reflectance data to absorbance spectral data, applying standard normal

variate (SNV) and then taking its second derivative. The second derivative of absorbance data was calculated using the gap-segment method with gap size and segment size equal to 7. The second derivative removes the baseline offset and resolves nearby peaks inaddition to sharpening of spectral features (Unscrambler version 9.0). These pre-treatments help in removing any scattering effects.

Non destructive classification of canola was done by applying PCA on spectral data and then applying linear discriminant analysis (LDA) on the principal components (PCs). PCA involves decomposition of X-variable matrix into uncorrelated variables, known as PCs. PCs are the uncorrelated variables and are described as the linear combination of all the predicting variables (wavelengths). The first PC accounts for maximum variability in the dataset and the subsequent PCs account for remaining variability in the dataset. The algorithm used for calculating the PCs was singular value decomposition (SVD). The wavelengths showing absolute maximum weight in each PC are considered as important wavelengths. The PCs enhances the spectral features and reduces dimensionality of the data.

Results and Discussion

The spectral data for the respective samples of canola seeds were averaged and mean spectral values of each sample were used in PCA. Figure 4.2 shows the average spectral data for each sample type.

Data were first classified broadly into two categories/classes on the basis of percent erucic acid present, canola (Sample 1) and not canola (Samples 2, 3, 4, 5, 6 and 7). To

reduce the dimensionality of the dataset principal component analysis (PCA) was applied on the spectral values. After applying PCA some data points were identified as outliers and were removed. The first 5 PCs in the model, formed after removal of outliers, accounted for 96% of the variation in the data with PC-1 taking into account 70% of the variation in the data. This means that maximum information of the variables can be obtained by using the first 5 PCs. These PCs were used as inputs to a LDA model. Unscrambler version 9.0 was used to perform the analysis. The data were divided into training and prediction datasets equally. Each dataset had 59 samples. The accuracy rate of 100% was achieved and the results obtained are shown in Table 4.3.

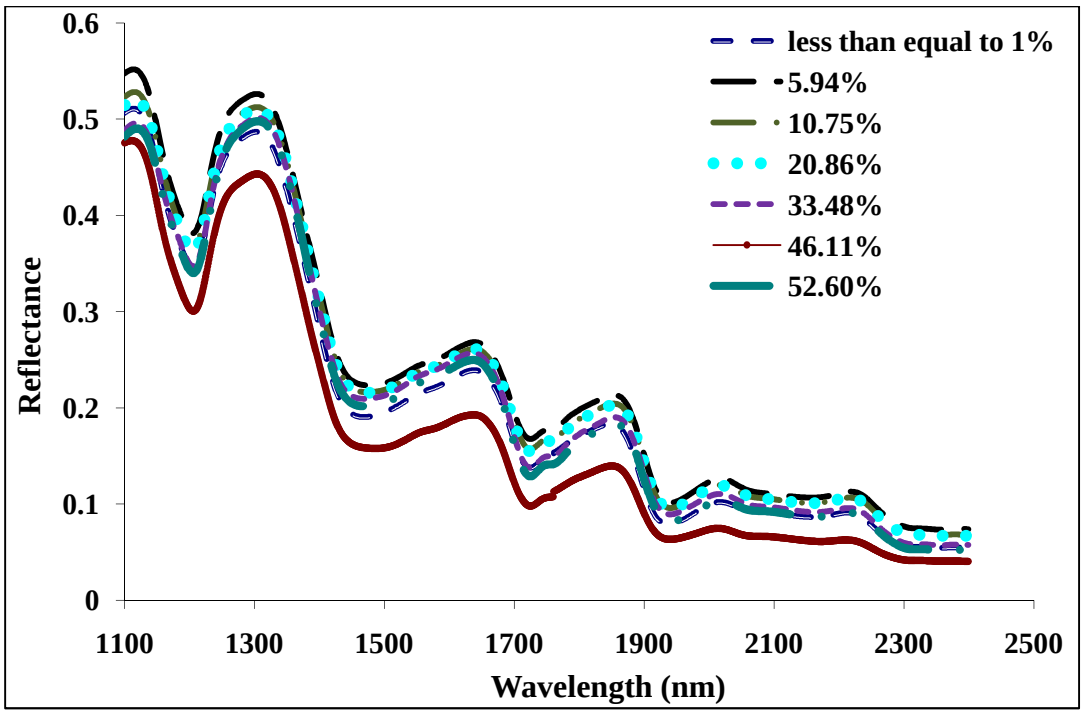


Figure 4.2. Average reflectance spectra in NIR range showing different levels of erucic acid in canola samples

To test the sensitivity of the method another test was carried out by categorizing samples within Sample 1. This was done to test if samples could be differentiated if they had erucic acid less than 1%. The dataset was categorized into six classes depending on the percent erucic acid present (Table 4.4). A PCA model was obtained after removing the outliers. The PC-1 accounted for 60% of the variation of the dataset. The first five PCs accounted for 92% of the variation in the dataset with less than 1% erucic acid. The dataset was divided into training (59 samples) and test datasets (53 samples). Table 4.4 show the results obtained after performing LDA on the training and test datasets. In the calibration model, one sample in class 2 was misclassified into class 1 and rest all the classes showed 100% accuracy. Class 1 and Class 2 had erucic acid 0 and 0.03% respectively. Since there is small difference between the two classes, it can be considered as a possible reason for the misclassification of the sample. An accuracy rate of 100% was achieved when the model was tested with the test dataset.

Table 4.3. Results obtained when samples were classified as canola and not canola using Linear Discriminant Analysis

Class	Calibration			Prediction		
	Number of Samples*	False Number**	Accuracy (%)	Number of Samples*	False Number**	Accuracy (%)
Canola	39	0	100	39	0	100
Not Canola	20	0	100	20	0	100

*Number of Samples: Total number of samples tested for calibration and prediction for both canola and not canola type; ** False Number: Misclassified samples

Based on the loadings of the PCA model, obtained after analysis on the whole data set (samples 1-6), important wavebands were identified. Wavebands at 1690, 1720, 1756, 1769 and 1860 nm were considered potential wavelengths responding to erucic acid. The identified wavelengths were then tested for their response to erucic acid variation using

multiple linear regression analysis. Spectral absorbance data corresponding to these wavelengths was used to develop a multivariate regression model. The results indicated the R-square value of 0.94 with standard error of prediction (SEP) equal to 5.26. Velasco et al (1997) in the study on intact Ethiopian-mustard seeds identified six regions, 1140-1240, 1350-1400, 1650-1800, 1880-1920, 2140-2200 and 2240-2380 nm , associated with fatty acid absorption. The results obtained from the PCA indicated region between 1690-1860 nm responding to change in erucic acid. The results took into account large variation in dataset (0-52.6 mg/g of seed sample) and different *Brassica* oilseed species (canola, rapeseed and mustard). The important identified regions overlapped with the research done on Ethiopian-mustard. The high R-square value shows the accuracy of the model. NIR spectroscopy can be considered as a potential method to estimate erucic acid and to distinguish between different varieties of *Brassica* family.

Table 4.4. Results obtained when the Sample 1 was classified into different classes depending upon erucic acid content using Linear Discriminant Analysis

Class	Erucic Acid Content (%)	Calibration			Prediction		
		Number of Samples	False Number	Accuracy (%)	Number of Samples	False Number	Accuracy (%)
1	0	10	0	100	9	0	100
2	0.03	7	1	87.5	9	0	100
3	0.08	19	0	100	19	0	100
4	0.22	5	0	100	4	0	100
5	0.72	6	0	100	2	0	100
6	1.02	10	0	100	10	0	100

*Number of Samples: Total number of samples tested for calibration and prediction for both canola and not canola type; ** False Number: Misclassified samples

Conclusions

A non-destructive approach to categorize seeds of *Brassica* family into canola and not canola was tested using NIRS. The LDA model was able to predict the samples into their respective categories with 100% accuracy. When samples with less than 1% erucic acid were categorized into their respective six categories using LDA, one sample in Class 2 was misclassified into Class 1 in the calibration model. However, when tested with a test dataset this model gave 100% accuracy. Also potential wavelengths, 1690,1720,1756,1769 and 1860 nm, responding to different levels of erucic acid were identified. The absorbance data corresponding to these wavelengths was used to estimate erucic acid with the R-square value of 0.94 and SEP of 5.26. It can be concluded that NIRS can be used for non-destructive classification of canola and other similar seeds in *Brassica* family with high level of accuracy.

References

Bewig, K., A. Clarke, C. Roberts, and N. Unklesbay. 1994. Discriminant analysis of vegetable oils by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society. 71(2):195-200.

Daun J.K. and P.C. Williams.1997. Near infrared analysis of oilseeds- current status and future directions. New techniques and applications in lipid analysis. AOCS Press, Champaign.266-282.

Greenwood C.F., J.A. Allen, A.S.Leong, T.N. Pallot, T.M. Golder and T. Golebiowski. An investigation of the stability of Nirs calibrations for the analysis of oil content in

whole seed canola. 10th International Rapeseed Congress, Canberra, Australia. http://regional.org.au/au/gcirc/1/572b.htm?. Accessed on 7 July 2010.

Hutton C.W., J.R. Garbow, and T.R. Hayes.1999.Nondestructive NMR determination of oil composition in transformed canola seeds. Lipids. 34 (12): 1339-1346.

Kanamangala, R.V., N.O. Maness ,M.W. Smith, G.H. Brusewitz, S. Knight and B. Chinta. 1999. Reduced lipid pecans:chemical alteration and implications for quality maintenance during storage . J. Amer. Soc. Hort. Sci. 124 :389-398.

Kumar, S., J. Chauhan, and A. Kumar. 2010. Screening for erucic acid and glucosinolate content in rapeseed-mustard seeds using near infrared reflectance spectroscopy. Journal of Food Science and Technology. 47(6):690-692.

Pallot T.N., A.S. Leong, J.A. Allen, T.M.Golder, C.F. Greenwood, T. Golebiowski.1999. Precision of fatty acid analysed using near infrared spectroscopy of whole seed brassicas.10th International Rapeseed Congress, Canberra, Australia. Accessed at : http://regional.org.au/au/gcirc/1/573.htm?. Accessed on 7 July 2010.

Pérez-Vich, B., L. Velasco, and J. Fernández-Martínez. 1998. Determination of seed oil content and fatty acid composition in sunflower through the analysis of intact seeds, husked seeds, meal and oil by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society. 75(5):547-555.

Sato T., I. Uezono, T. Morishita and T. Tetsuka.1998.Nondestructive estimation of fatty acid composition in seeds of Brassica napus L. by near-infrared spectroscopy. Journal of the American Oil Chemists Society. 75(12):1877-1881.

Sato T., Y. Takahata, T. Noda, T. Yanagisawa, T. Morishita, and S. Sakai.1995. Nondestructive determination of fatty acid composition of husked sunflower (Helianthus annua L.) seeds by near-infrared spectroscopy. Journal of the American Oil Chemists' Society.72 (10):1177-1183.

Siemens B.J. and James K. Daun. 2005. Determination of the fatty acid composition of canola, flax and solin by near-infrared spectroscopy. Journal of the American Oil Chemists Society. 82(3):153-157.

Sundaram, J., C. Kandala, and C. Butts. 2009. Application of near infrared spectroscopy to peanut grading and quality analysis: overview. Sensing and Instrumentation for Food Quality and Safety. 3(3):156-164.

Tigabu, M., P. C. Oden, and D. Lindgren. 2005. Identification of seed sources and parents of Pinus sylvestris L. using visible–near infrared reflectance spectra and multivariate analysis. Trees - Structure and Function. 19(4):468-476.

Velasco, L., J.M. Fernandez-Martinez, and A. De Haro.1997. Determination of the fatty acid composition of the oil in intact-seed mustard by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists Society. 74(12):1595-1602.

Velasco, L. and H.C. Becker.1998. Estimating the fatty acid composition of the oil content in intact-seed rapeseed (Brassica napus L.) by near-infrared reflectance spectroscopy. Euphytica, Kluwer Academic Publishers. 101:221-230.

Velasco, L., F. Goffman, and H. Becker. 1999. Development of calibration equations to predict oil content and fatty acid composition in brassicaceae germplasm by near-

infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society. 76(1):25-30.

Zhu, S., J. Liang, and L. Yan. 2011. Study on Rapid Identification Methods of Transgenic Rapeseed Oil Based on Near Infrared Spectroscopy. In Computer and Computing Technologies in Agriculture IV, 633-640. D. Li, Y. Liu, and Y. Chen, eds: Springer Boston.

Zou, Q., H. Fang, F. Liu, W. Kong, and Y. He. 2011. Comparative Study of Distance Discriminant Analysis and Bp Neural Network for Identification of Rapeseed Cultivars Using Visible/Near Infrared Spectra. In Computer and Computing Technologies in Agriculture IV, 124-133. D. Li, Y. Liu, and Y. Chen, eds: Springer Boston.

CHAPTER V

NON DESTRUCTIVE ESTIMATION OF FREE FATTY ACID CONTENT AND PEROXIDE VALUE USING NIR SPECTROSCOPY IN CANOLA SEED

Abstract: Presence of free fatty acids (FFA) and peroxide value (PV) are considered as indicators for the degree of rancidity of oilseeds. The estimation of FFA and PV involves wet chemical laboratory analysis which requires destruction of sample and is a time consuming process. Non destructive estimation of FFA and PV using near-infrared spectroscopy (NIRS) in canola was outlined as the main objective of this study. Both FFA content and PV were successfully estimated using a partial least square (PLS) regression model. Important wavebands in the NIR range were also identified. Wavebands around 1685, 1850, 2055, 2092 and 2242 nm were used for estimation of FFA. The wavelengths around 1158, 1209, 1384, 1341, 1417, 1475, 2128, 2250 and 2303 nm were considered as potential wavebands for the estimation of PV.

Review of Literature

Rancidity causes musty or sour odor in the seeds. This kind of odor decreases the grade of canola to US sample grade. These odors are mainly present due to the presence of ketones and aldehydes as the oil in seeds undergoes oxidation (Caddick, SGRL) or in other words rancidity. Rancidity is mainly caused by hydrolysis of triglycerides to free

fatty acids (FFA) and then oxidation of FFA into ketones and aldehydes. It can be categorized into hydrolytic rancidity and oxidative rancidity. Presence of high amount of FFA can be used as an indicator of hydrolytic rancidity and oxidative rancidity is measured by determining the peroxide value (PV). PV measures the extent of oxidation by determining the amount of peroxides present. Therefore, percent FFA content or peroxide value (PV) can also be used to determine if the seed is rancid or not. According to Canadian General Standards Board requirements for Canola oil, FFA content in crude oil should not be more than 1% (by mass as oleic acid). In refined, bleached and deodorized oil maximum permissible FFA content and PV is 0.05% and 2 meq/kg respectively (Canola: Standards and Regulations, CGSB 1987). According to Canadian Grain Commission top grade canola seed usually has less than 0.7% free fatty acids. U.S has not published any similar food grade specifications in regard to canola oil.

U.S. graders primarily smell a sample of seeds to grade canola on basis of odor. But these odors can be easily eliminated or suppressed by a producer by using fumigants or other odor suppressants. If seed sample is sent to laboratory for testing using wet chemical laboratory methods, then it is an expensive and time consuming process in addition, it involves sample destruction. A more cost effective efficient method could be the use near infrared spectroscopy (NIRS). NIRS can simultaneously measure different oil parameters like oil, water and protein content, fatty acid composition, percent FFA and peroxide value (PV) etc.

Near-infrared radiation covers the region from 700 – 2500 nm in the electromagnetic spectrum. It mainly responds to vibration energy levels of a molecule. The type of vibrations of a particular functional group determines absorption spectrum in the infrared

region of the electromagnetic spectrum. Functional groups O-H, C-H, C-OOH, C=O, aromatic C-H groups and N-H are mainly identified in the NIR region. Thus food quality, which is affected by amount of moisture content, organic acids, food oxidative products and sugars present in it, can be measured using NIR spectroscopy (Sundaram et al., 2009). NIR spectroscopy can be used as a non destructive technique for oilseed analysis (Daun and Williams, 1997). It has been successfully used to estimate the oil content in whole seed canola (Greenwood *et al*, 1999). The calibration model was based on the samples analysed from 1995 to 1997. Stability of the model still needed to be evaluated to predict oil content with future years of canola samples. Sato et al (1998) had successfully evaluated the feasibility of using NIR spectroscopy as a non-destructive method for estimation of fatty acid composition of rapeseed. Velasco and Becker (1998) were able to develop highly accurate calibrations for oleic, linoleic, linolenic and erucic acid for rapeseed using 3 g sample. In another study, Velasco *et al* (1997) identified six spectral regions: 1140-1240, 1350-1400, 1650-1800, 1880-1920, 2140-2200 and 2240-2380 nm to estimate the fatty acid composition of oil in intact seed samples of Ethiopian mustard. Velasco *et al* (1999) conducted a study to develop calibration equations to estimate seed oil content and concentration of major fatty acids for intact seeds samples of Brassicaceae family. It was concluded that only oil content and concentrations of $C_{18:1}$, $C_{18:3}$ and $C_{22:1}$ can be estimated reliably by using NIR spectroscopy. Calibrations for non destructive estimation of erucic acid and glucosinate content in rapeseed and mustard seeds have been developed by using partial least square regression on data obtained from Fourier transform near infrared reflectance spectroscopy (FT-NIR). A close relationship

was observed between reference method and predicted values from the spectral data from 7502.1 to 5444.6 cm^{-1} (Kumar et al., 2010).

NIR spectroscopy was also applied to determine the fatty acid (FA) composition of husked sunflower (Sato *et al* 1995). In one of the studies it was observed that the accuracy of the calibration model based on NIR spectroscopy to estimate oil content and fatty acid decreased when intact (non husked) sunflower seed was used as compared to the accuracy obtained from oil, meal or husked seed (Pérez-Vich et al., 1998). Seimens and Daun (2005) used partial least square (PLS) analysis to develop calibration for determining fatty acid composition of canola, flax and solin. It was concluded that NIR spectroscopy showed potential in estimating linolenic acid, oleic acid and total saturated FA. Pallot *et al* (1999) also developed calibrations for eight major fatty acids (C16:0, C18:0, C18:1, C18:2, C18:3, C20:0, C20:1, C22:1) using full NIR spectral information instead of selected wavelengths, for each separate group of samples (canola, high erucic acid rapeseed, canola quality and condiment mustard samples). Nuclear magnetic resonance (NMR) spectroscopy has also been evaluated as a non-destructive technique to develop seed oil calibrations (Hutton, Garbow and Hayes; 1999).

Apart from non-destructive estimation of fatty acid composition, NIR spectroscopy has also been used to discriminate between rapeseed cultivars (Zou et al., 2011). This was done using distance discriminant analysis (DDA) and back propagation neural network (BPNN). DDA and BPNN were applied on the first six principal components (PCs) obtained after applying principal component analysis (PCA) on the spectral data. In this study 225 samples from 5 kinds of rapeseed were used. 100% prediction accuracy was achieved by both the methods but it was concluded that DDA was a better approach over

BPNN since it was convenient and more intuitive. Another study showed an overall correct identification rate of 96.34% when transgenic rapeseed oil was distinguished from non transgenic rapeseed oil (Zhu et al., 2011). This study used PCA in conjunction with discriminant partial least squares (DPLS). PCA score plot showed that both oil types can be distinguished from each other roughly. A similar study on *Pinus sylvestris* L. showed that partial least squares discriminant analysis (PLS-DA) showed a 100% classification accuracy when seeds were identified on the basis of seed sources and parents using visible (VIS) and NIR spectroscopy (Tigabu et al., 2005). Discriminant analysis on NIR data has also been used classify four vegetables oil types (cotton-seed, peanut, soybean and canola). Four wavelengths, 1704,1802,1816 and 2110 nm were identified and were used to build the model (Bewig et al., 1994).

Fourier Transformed infrared (FTIR) spectroscopy has been evaluated to estimate FFA in fats and oils. The amount of FFA ranged between 0.2-8% and the results obtained were comparable to that of America Oil Chemists' Society (AOCS) reference titration method (Ismail *et al*, 1993). FTIR spectroscopy has also been applied to develop successful calibration models to estimate PV of vegetable oils (Voort *et al*, 1994). Moschner and Korell (2006) developed an efficient and a cost effective method to estimate FFA content in sunflower seeds using NIRS. An R-square value of 0.94 with a standard error of prediction equal to 0.2 was achieved using this method. The model was developed using modified partial least square algorithm. NIR spectral analysis to estimate FFA in crude palm oil, refined-bleached-deodorized (RBD) palm olein and RBD palm oil was done using multiple linear regression (Man *et al.,* 1998). The wavelengths combinations including 1882, 2010 and 2040 nm were identified as important combinations. The

method was efficient with time required for the analysis less than 5 min and it accuracy was considered good for raw material quality control. The important wavelengths were selected using a stepwise multiple linear regression in spectral range of 1850-2050 nm since C=O stretching bands lie in this region. NIRS can be used to determine the food quality since it can estimate different parameters simultaneously (Gerde *et al.*, 2007). It can be used to measure the hydrolytic lipid degradation of fish oil by the fish meal industry (Cozzolino *et al.*, 2005). The PLS regression model showed good accuracy for estimating FFA and moisture but the poor calibration model accuracy was obtained for models developed to estimate PV and anisdine value (AV). A study conducted on different edible oils, olive oil, maize, seed and sunflower showed that less than 30 s is required for direct acidity quantification and screening of peroxide index using NIRS technology (Armenta *et al.*, 2007). A study conducted by Yildiz *et al* (2003) showed high correlations between NIR and PeroxySafe^{TM} kit data to estimate PV in the soybean oil. Moh *et al* (1999) developed PLS calibration model to estimate PV of crude palm oil (CPO) using NIR spectral data in the spectral region from 1350 -1480 nm. This region was found to have high correlation between PV and NIR spectra. A study conducted on corn and soybean oils showed that the PLS regression model formed using first derivative spectra showed a correlation coefficient of 0.99 between titration and NIR method (Yildiz *et al.*, 2002). Spectral region around 2068 nm was associated with the hyper-oxides formed during oil oxidation. NIRS can be used to measure the oxidation of soybean oil (Yildiz *et al.*, 2001).

NIRS has proven to be fast, efficient, environmentally friendly and an accurate method to estimate both PV and FFA content in oils (Armenta et al., 2007; Gerde et al., 2007; Moh

et al., 1999). As part of this study, NIR spectroscopy is being evaluated for non destructive estimation FFA and PV in canola seeds using NIRS.

Material and Methods

Samples

A total of 140 samples of eight types were analyzed. The variation in FFA and PV ranged from 0.7 to 1.8% and 1.49 to 7.77 meq /kg of sample (Table 5.1). Samples1, 2, 3 and 8 were part of canola obtained from a local farmer in 2007. Sample 1 was tested when it was received that is on Day 0. Sample 2 was stored in an oven at 45 degree C for a period of 25 days so as to change it chemical properties. Sample 3 was obtained by storing under natural hot and humid conditions in a grain bin for a period of 60 days. Samples 4, 5, 6 and 7 were obtained from a local farmer in 2009. Sample 4 was tested at day 0 when it was received. Sample 5 was kept under natural hot and humid conditions for 30 days in a grain bin. Since weather conditions changed therefore rest of canola sample was shifted to an environment chamber where samples 6 and 7 were stored at 40 degree C for 7 and 8 months respectively before getting tested. Sample 8 was part of 2007 canola stored over a period of two years in a walk in cooler at Biosystems and Agricultural Engineering Labs, Oklahoma State University. These samples were also sent to Enid Grain Inspection Company, Enid, OK for testing these samples for musty/sour odor.

Table 5.1. FFA and PV values of canola samples

Sample #	Treatment	FFA (%)	PV (meq. peroxide/kg sample)	Number of Samples
1*	Day 0	0.77	4.24	10
2*	Stored at 45 degree C for 25 days	1.85	6.18	10
3*	Stored under natural hot and humid conditions for 60 days	1.3	7.77	20
4**	Day 0	1.66	1.49	20
5**	Stored under natural hot and humid conditions for 30 days	1.46	2.11	20
6**	Stored at 40 degree C for 7 months	1.45	5.03	20
7**	Stored at 40 degree C for 8 months	1.63	3.73	20
8*	Stored in walk in cooler for 2 years	1.28	5.54	20

*Canola from 2007 season, ** Canola from 2009 season

Oil extraction

20 g samples of each sample type were ground using a 2-L Waring blender at low speeds for 2 minutes. The blender was kept ‘on’ for 20 seconds and then it was turned ‘off’. The seed was ground at 20 second interval for a total of six times such that total time the sample was ground was 2 minutes. The sample was ground for short intervals to avoid any heating of the sample. The sample was ground to pass through a 1 mm sieve. Oil was extracted from the seed samples using ether extraction with a rotary evaporator. Three replications were done for each type of sample. The oil extracted, was used to determine its percent FFA content and PV.

Percent FFA determination

The procedure used was based on AOAC official method 940.28 (1999). From the extracted oil, 1.41 g of oil was weighed into a test tube. In another test tube 10 mL of

ethanol was neutralized by adding 0.4 mL of phenolphthalein solution and 0.1N NaOH. Phenolphthalein was used as an indicator. A faint permanent pink color indicated neutralization of ethanol. This solution was then added to the oil. This mixture of oil and neutralized alcohol was titrated with 0.025 N NaOH with vigorous shaking until a permanent faint pink color appeared and persisted for greater than 1 min. The amount, in mL, of 0.025N NaOH used in titration corresponded to percent free fatty acids expressed as oleic acid (equation 1).

$$\text{FFA (\%)} = (V)\,(N)\,(28.2)/\text{weight of sample} \quad (1)$$

where V= volume of 0.025 N NaOH used for titration of sample
N = normality of NaOH used for titration, in this example it is 0.025 N

PV determination

The procedure used was based on AOCS Official Method Cd 8-53 (same as AOAC Official Method 965.33). 0.5 g of oil was weighed into a 250 ml glass-stoppered Erlenmeyer flask and 3 ml of acetic acid: chloroform (3:2) is added to it. Flask is then swirled until oil is dissolved in solvent and 0.05 ml of saturated potassium iodide is added to it. Let the mixture stand for 1 minute and then add 3 ml of de-ionized water. The solution was then titrated with 0.001 N sodium thiosulphate until yellow color has almost disappeared. At this point 0.05 ml of starch indicator solution (1%) was added to the flask and the mixture was then titrated again by adding sodium thiosulphate drop wise, until blue color had just disappeared. Constant and vigorous shaking of flask should be done while doing this titration. PV is estimated using equation 2.

PV (meq. peroxide/kg sample) = (S - B)(N)(1000)/weight of sample (2)

where S= volume of sodium thiosulphate for titration of sample
B= volume of sodium thiosulphate for titration of blank
N=normality of sodium thiosulphate

Spectral Data Acquisition

Spectral data, within the spectral range from 350nm to 2500nm, were collected using a FieldSpec Pro spectrometer (ASD Inc., CO). It has a spectral resolution of 3nm at 700 nm with a sampling interval of 1.4 nm between 350 -1050 nm. The spectrometer has a spectral resolution of 10 nm at 1400 and 2100 nm with a sampling interval of 2 nm between 1000- 2500 nm. The field of view was 25 degrees and reflectance was recorded from a height of 2.9 cm. Since the reflectance data was not taken at 90 degree angle the scattering of wavelengths was taken into account by rotating the sample using a turn table (Figure 5.1). The probe was recalibrated using a white standard having 99 percent reflectance (Spectralon Target, Labsphere, NH) before recording data for each sample. Due to high noise level at the ends of the spectra the regions from 300-400 nm and 2400-2500 nm were not used in the analysis. Data within the NIR range of 1100 nm to 2400 nm was used for further analysis.

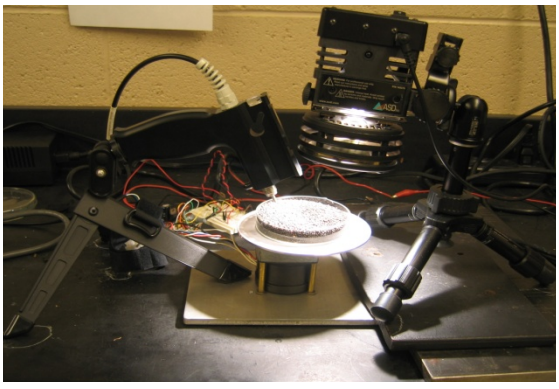


Figure 5.1. Set up for spectral data acquisition

Data Processing

Before analyzing the spectral data, the data was pre-processed using Unscrambler 10.0.1 (CAMO Software Inc., Woodbridge, NJ, USA). Log (1/R), standard normal variate (SNV) and de-trending mathematical transformations were applied on spectral data used to estimate FFA and PV. Second order derivative was performed using gap segment method with gap and segment size equal to 9. Partial least squares regression method was used to estimate FFA and PV. The cross validation of the model was done using 17 random segments from the data with 76 samples per segment. The accuracy of the model was described by the standard error prediction (SEP) and multiple coefficient of determination (R-square).

Results and Discussion

When samples were sent to Enid Grain Inspection Company to test the presence of musty/sour odor in the seeds, only sample 8, was graded as sample grade canola with musty/sour odor. The wet chemical laboratory tests indicated that the FFA and PV of samples 2, 3 and 6 were either greater or approximately equal to that of sample 8, but these samples were categorized as US grade 1 canola by the graders. This shows that sample odor is a subjective method to grade canola. Past research has shown that NIRS is a more reliable and accurate technique for the estimation of FFA and PV.

Figure 5.2 shows the average spectral values of different samples in NIR region. PLS regression model was used to estimate percent FFA content in canola seed. Wavelengths in NIR range were considered as independent variables (X) and the percent FFA content

obtained after wet chemical laboratory analysis as dependent variable (Y). A random cross validation of the samples was done with 17 segments and each segment contained 76 samples. Estimation of percent FFA was done by applying only SNV and second derivative transformations. The PLS model was formed using six factors explaining 92.6 % of the variation in the samples. The model yielded the R-square value of 0.92 for calibration data and 0.89 for validation data. SEP of 0.15 was obtained. Out of 140 samples, 18 samples were considered as outliers and were removed from the analysis.

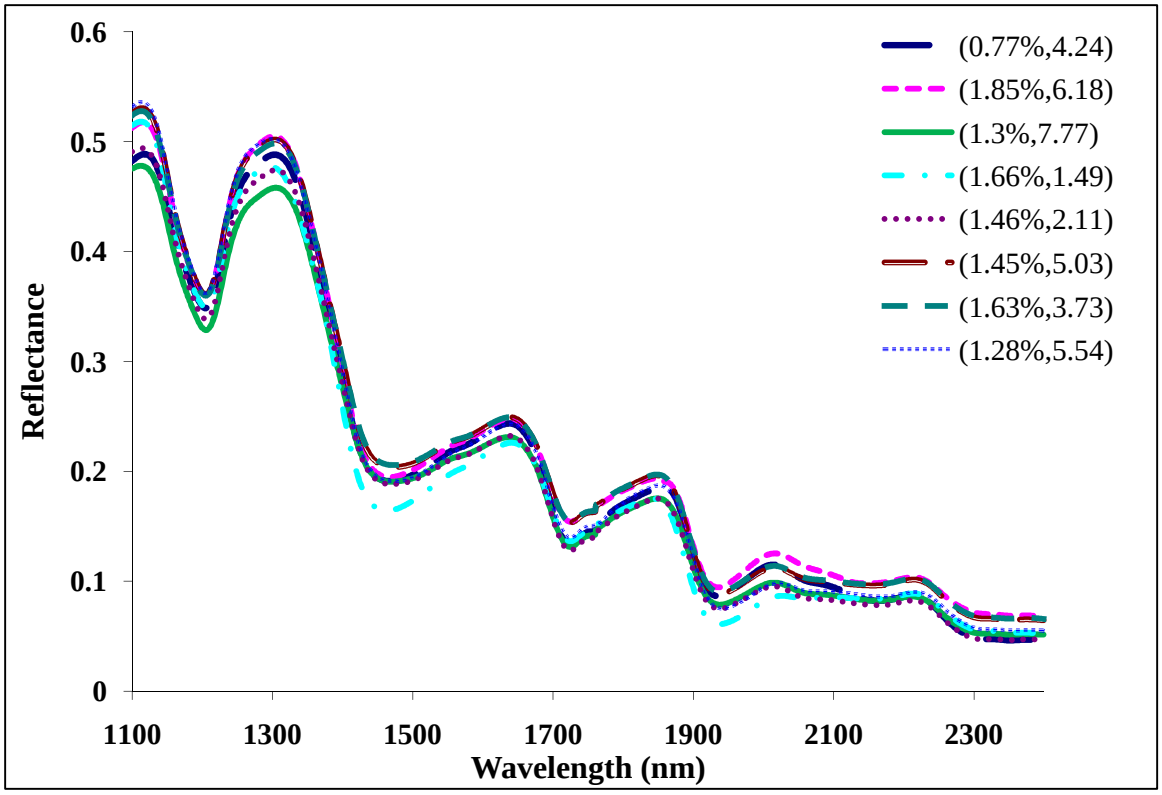


Figure 5.2. NIR spectra for each sample having different FFA content and PV (FFA%, PV)

The estimation of PV was done by applying SNV, de-trending and second derivative transformations on the NIR spectral data. The model was formed using 7 factors

explaining 94% variation in the data. The PLS model yielded the R-square value of 0.94 for calibration data and 0.88 for validation data. SEP of 1.56 was observed for this model. This model performed better than the FTIR model developed by the Voort et al (1994) for vegetable oils (overall SEP =2.66 PV).

Important wavelengths were also identified using the loadings plot of both the models to estimate percent FFA and PV. The wavebands at 1685, 1850, 2055, 2092 and 2242 nm were considered as important wavebands to estimate percent FFA. A multiple linear regression (MLR) model generated using non-transformed absorbance data across these wavebands had the R-square value of 0.89 (calibration) and 0.88 (validation). The value of SEP for this model was 0.09. FFA is characterized by carboxylic acids, thus C=O stretching band in the NIR region (1850-2050 nm) is considered important for their estimation (Man and Moh, 1998). This study on palm oil identified 1882, 2010 and 2040 nm as best wavelength combination for FFA determination. These wavelengths were close to the wavelengths identified in the current study. The difference in the selected wavelengths could be due to different FFA composition and degree of degradation of both the oils.

Important wavebands identified for estimating PV were: 1158, 1209, 1384, 1341, 1417, 1475, 2128, 2250 and 2303 nm. The MLR model formed using these wavebands had the R-square value of 0.90 (calibration) and 0.89 (validation). The model had a SEP of 0.68. Yildiz et. al (2001) in a study on oxidation of vegetable oils identified important wavelengths :1400,1746, 2036 and 2070 nm to estimate lipid peroxide when first derivative was applied on 1mm path length spectra. When data were analyzed for 2 mm path length 2068, 2016, 1612 and 1242 nm were identified as key wavelengths. On

comparing the results obtained by Yildiz et. al (2001) with the results obtained in the current study it is observed that region around 1400 and 2100 nm holds important information. Takamura et al (1995) had also reported 2084 nm as an important wavelength for determining lipid oxidation in edible oils. The single wavelength approach showcases some problems due to interferences caused by other similar functional groups having hydroperoxide band (alcohols, water). This would cause errors in accurate measurement of PV. A multispectral approach would overcome this problem.

In the current study, the models formed by MLR showed lesser SEP values than the models yielded by PLS analysis using the whole spectral information. The results indicate that FFA and PV can be estimated using NIR spectral data.

Conclusion

This study demonstrates the non destructive estimation of FFA content and PV using NIR spectroscopy. The developed PLS models showed the R-square values of 0.94 and 0.90 for FFA content and PV respectively. The accuracy of the model was determined using the SEP. The SEP of 0.15 and 1.56 was observed for FFA and PV models respectively. Important wavelengths, 1685, 1850, 2055, 2092 and 2242 nm were identified for the estimation of FFA. The wavelengths around 1158, 1209, 1384, 1341, 1417, 1475, 2128, 2250 and 2303 nm were considered as potential wavebands for the estimation of PV. The developed method is non-destructive method and consumes less time. This would aid in faster and on-site grading of samples.

References

AOCS. 1997. The Official Methods and Recommended Practices of the American Oil Chemists' Society, 4th edition. AOCS Official Method Cd 8-53: Peroxide Value – Acetic Acid-Chloroform Method, Re-approved 1997.

AOAC Official Method 940.28 (1999): Fatty acids (free) in crude and refined oils. Vol. 2. Cunniff P. (ed.) 16th Ed. Official Methods of Analysis of AOAC international. AOAC International, Gaithersburg, Chapter 41: 10

Armenta, S., S. Garrigues, and M. de la Guardia. 2007. Determination of edible oil parameters by near infrared spectrometry. Analytica Chimica Acta 596(2):330-337.

Bewig, K., A. Clarke, C. Roberts, and N. Unklesbay. 1994. Discriminant analysis of vegetable oils by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society. 71(2):195-200.

Caddick Len. Why are oilseeds different to store. CSIRO Stored Grain Research Laboratory.

Canadian Grain Commision.Oilseeds methods and tests used to measure quality. http://www.grainscanada.gc.ca/oilseeds-oleagineux/method-methode/omtm-mmao-eng.htm. Accessed on 27 April 2009.

Canola: Standards and Regulations. Canadian General Standards Board requirements for Canola Oil. 1987. Available at: http://www.canolacouncil.org/uploads/Standards1-2.pdf. Accessed on: 19 February 2011.

Cozzolino, D., I. Murray, A. Chree, and J. Scaife. 2005. Multivariate determination of free fatty acids and moisture in fish oils by partial least-squares regression and near-infrared spectroscopy. LWT - Food Science and Technology. 38(8):821-828.

Daun J.K. and P.C. Williams.1997. Near infrared analysis of oilseeds- current status and future directions. New techniques and applications in lipid analysis. AOCS Press, Champaign. 266-282.

Gerde, J., C. Hardy, C. Hurburgh, and P. White. 2007. Rapid Determination of Degradation in Frying Oils with Near-Infrared Spectroscopy. Journal of the American Oil Chemists' Society. 84(6):519-522.

Greenwood C.F., J.A. Allen, A.S.Leong, T.N. Pallot, T.M. Golder and T. Golebiowski. An investigation of the stability of Nirs calibrations for the analysis of oil content in whole seed canola. 10th International Rapeseed Congress, Canberra, Australia. http://regional.org.au/au/gcirc/1/572b.htm?. Accessed on 7 July 2010.

Hutton C.W., J.R. Garbow, and T.R. Hayes.1999.Nondestructive NMR determination of oil composition in transformed canola seeds. Lipids. 34 (12): 1339-1346.

Ismail A.A., F.R. van de Voort, G. Emo and J. Sedman.1993. Rapid quantitative determination of free fatty acids in fats and oils by Fourier transform infrared spectroscopy. Journal of the American Oil Chemists Society. 70(4):335-341.

Kumar, S., J. Chauhan, and A. Kumar. 2010. Screening for erucic acid and glucosinolate content in rapeseed-mustard seeds using near infrared reflectance spectroscopy. Journal of Food Science and Technology. 47(6):690-692.

Man, Y., and M. Moh. 1998. Determination of free fatty acids in palm oil by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society. 75(5):557-562.

Moh, M., Y. Che Man, F. van de Voort, and W. Abdullah. 1999. Determination of peroxide value in thermally oxidized crude palm oil by near infrared spectroscopy. Journal of the American Oil Chemists' Society. 76(1):19-23.

Moschner, C. R., and B. Biskupek-Korell. 2006. Estimating the content of free fatty acids in high-oleic sunflower seeds by near-infrared spectroscopy. European Journal of Lipid Science and Technology. 108(7):606-613.

Pallot T.N., A.S. Leong, J.A. Allen, T.M.Golder, C.F. Greenwood, T. Golebiowski.1999. Precision of fatty acid analysed using near infrared spectroscopy of whole seed brassicas.10th International Rapeseed Congress, Canberra, Australia. http://regional.org.au/au/gcirc/1/573.htm?. Accessed on 7 July 2010.

Pérez-Vich, B., L. Velasco, and J. Fernández-Martínez. 1998. Determination of seed oil content and fatty acid composition in sunflower through the analysis of intact seeds, husked seeds, meal and oil by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society. 75(5):547-555.

Sato T., I. Uezono, T. Morishita and T. Tetsuka.1998.Nondestructive estimation of fatty acid composition in seeds of Brassica napus L. by near-infrared spectroscopy. Journal of the American Oil Chemists Society. 75(12):1877-1881.

Sato T., Y. Takahata, T. Noda, T. Yanagisawa, T. Morishita, and S. Sakai.1995. Nondestructive determination of fatty acid composition of husked sunflower (Helianthus annua L.) seeds by near-infrared spectroscopy. Journal of the American Oil Chemists' Society.72 (10):1177-1183.

Siemens B.J. and James K. Daun.2005.Determination of the fatty acid composition of canola, flax and solin by near-infrared spectroscopy. Journal of the American Oil Chemists Society. 82(3):153-157.

Sundaram, J., C. Kandala, and C. Butts. 2009. Application of near infrared spectroscopy to peanut grading and quality analysis: overview. Sensing and Instrumentation for Food Quality and Safety. 3(3):156-164.

Takamura, H., N. Hyakumoto, N. Endo , and T. Matoba,. 1995. Determination of lipid oxidation in edible oils by near-infrared spectroscopy. J. Near Infrared Spectrosc. 3:219–225.

Tigabu, M., P. C. Oden, and D. Lindgren. 2005. Identification of seed sources and parents of Pinus sylvestris L. using visible–near infrared reflectance spectra and multivariate analysis. Trees - Structure and Function. 19(4):468-476.

Velasco L, J.M. Fernandez-Martinez, and A. De Haro.1997. Determination of the fatty acid composition of the oil in intact-seed mustard by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists Society. 74(12):1595-1602.

Velasco l. and H.C. Becker.1998. Estimating the fatty acid composition of the oil content in intact-seed rapeseed (Brassica napus L.) by near-infrared reflectance spectroscopy.Euphytica, Kluwer Academic Publishers. 101:221-230.

Velasco, L., F. Goffman, and H. Becker. 1999. Development of calibration equations to predict oil content and fatty acid composition in brassicaceae germplasm by near-infrared reflectance spectroscopy. Journal of the American Oil Chemists' Society. 76(1):25-30.

Voort F.R. van de, A.A. Ismail, J. Sedman, J. Dubois and T. Nicodemo.1994. The determination of peroxide value by Fourier transform infrared spectroscopy. Journal of the American Oil Chemists Society. 71(9):921-926

Yildiz, G., R. Wehling, and S. Cuppett. 2001. Method for determining oxidation of vegetable oils by near-infrared spectroscopy. Journal of the American Oil Chemists' Society. 78(5):495-502.

Yildiz, G., R. Wehling, and S. Cuppett. 2002. Monitoring PV in corn and soybean oils by NIR spectroscopy. Journal of the American Oil Chemists' Society. 79(11):1085-1089.

Yildiz, G., R. Wehling, and S. Cuppett. 2003. Comparison of four analytical methods for the determination of peroxide value in oxidized soybean oils. Journal of the American Oil Chemists' Society. 80(2):103-107.

Zhu, S., J. Liang, and L. Yan. 2011. Study on Rapid Identification Methods of Transgenic Rapeseed Oil Based on Near Infrared Spectroscopy. In Computer and Computing

Technologies in Agriculture IV, 633-640. D. Li, Y. Liu, and Y. Chen, eds: Springer Boston.

Zou, Q., H. Fang, F. Liu, W. Kong, and Y. He. 2011. Comparative Study of Distance Discriminant Analysis and Bp Neural Network for Identification of Rapeseed Cultivars Using Visible/Near Infrared Spectra. In Computer and Computing Technologies in Agriculture IV, 124-133. D. Li, Y. Liu, and Y. Chen, eds: Springer Boston.

CHAPTER VI

CONCLUSIONS

This study focused on developing techniques that may assist in determining the quality of canola. The main focus of these techniques was to reduce the human error imparted while grading canola. Grading of canola is mainly done by visual inspection, smelling seed or by performing wet chemical analysis in a laboratory setting. Analyzing canola samples by wet chemical tests not only destroys the sample but is a time consuming and a tiring process. In addition, it requires specialized training to estimate the quality of a seed sample. The techniques discussed in the study will not only decrease time to analyze the sample but will also be more accurate. Canola seed samples were analyzed using a flat-bed scanner to estimate the amount of total dockage (dockage + conspicuous admixture) present in the sample. To accomplish this, different samples with varying degree of total dockage: 0.1, 0.3, 1, 1.34, 1.45, 2.5, 3.4, 4.3, 4.6 and 11.3%, were prepared and tested. Mean values of luminosity (L), red (R), green (G) and blue (B) were recorded for 20 sub-samples taken from each sample type. One way ANOVA applied on the L, R, G and B data of the samples showed that the sample with the most dockage was significantly different in all the domains. According to Tukey's HSD method only sample 10 with

11.2% of total dockage was significantly different from all the other samples. The discriminant analysis showed a misclassification rate of 14%. The sample with 2.5% of total dockage showed maximum error during classification of samples. This method showed some potential but was not precise and accurate enough to provide a grading tool.

Another objective of this study was to identify distinctly green (DG) and heat damaged seeds `HD). According to USDA standard guidelines, US grade 1 canola should not have more than 3% of the total damaged seeds (distinctly green seeds and heat treated seeds). DG seeds should not be more than 2% and HD seeds should not be more than 0.1%. Graders usually identify these seeds by crushing a sample of canola and inspecting the crushed seeds visually. The results from this study indicated 100 % classification accuracy for distinctly green seeds but some of the good and heat damaged seeds were misclassified. It was concluded that machine vision is a potential method for grading of canola into good, DG and HD seeds, on the basis of seed color.

Classification of canola seed from other high erucic acid seeds on the basis of erucic acid using near infrared spectroscopy (NIRS) was outlined as the third objective of this study. Traditionally amount of erucic acid is determined using gas chromatography, which is an expensive and a time consuming procedure. Researchers have been successful in determining various free fatty acids such as linoleic acid, linolenic acid and oleic acid, using NIR spectroscopy. Principal Component Analysis was applied on the pre-processed spectral data. This was done to reduce the dimensionality and to identify important wavebands in the spectra responding to change in erucic acid between the samples. A linear discriminant model was built using the principal components obtained. Samples

were classified broadly into seven classes (less than equal to 1.02 % to 52.6% erucic acid) with 100% accuracy. Sensitivity of the analysis was also tested by analyzing canola samples with different levels of erucic acid (0 to 1.02% erucic acid). Accuracy of 100% was achieved when a test dataset was tested on the calibration model. Wavebands at 1690,1720,1756,1769 and 1860 nm showed good response to change in erucic acid levels and thus, can be considered potential wavelengths to estimate erucic acid.

Non destructive estimation of free fatty acid (FFA) content and peroxide value (PV) using NIRS on canola was outlined as another objective of this study. Both FFA and PV are considered as indicators of the degree of rancidity. According to the Canadian General Standards Board requirements for Canola oil, FFA content in crude oil should not be more than 1% (by mass as oleic acid). In refined, bleached and deodorized oil maximum permissible FFA content and PV is 0.05% and 2 meq/kg, respectively (Canola: Standards and Regulations, CGSB 1987). The estimation of FFA and PV involves wet chemical laboratory analysis, which requires destruction of sample and is time consuming. Both FFA and PV were successfully estimated using a partial least square (PLS) regression model. Important wavebands in the NIR range were identified. Wavebands around 1685, 1850, 2055, 2092 and 2242 nm were used for the estimation of FFA. The wavelengths around 1158, 1209, 1384, 1341, 1417, 1475, 2128, 2250 and 2303 nm were considered as potential wavebands for the estimation of PV.

Future work in this area should include development of an electronic sensor system that will help a grader or an un-trained farmer to estimate the quality of canola seed. This sensor should include the identified wavelengths in this study for the estimation of erucic acid, FFA content and PV of a seed sample. The sensor systems developed would assist

the graders in accurately determining the grade of canola samples. The algorithm developed for the classification of good, DG and HD seeds though performed well, but to account for the variation in the seeds more data for DG seeds should be collected. Inclusion of more data of individual seed type would make the algorithm more robust.

APPPENDIX I

MATLAB CODE FOR CLASSIFICATION OF GOOD, DISTINCTLY GREEN AND HEAT DAMAGED CANOLA SEEDS

```
close all
clear all
clc
%reading the files
Im=imread('C:\Documents and Settings\geetika\My Documents\Dr
Jones\Canola grading\green-yellow grading\strip005', 'jpg');
figure(1);imshow(Im,[]);
r=Im(:,:,1);
g=Im(:,:,2);
b=Im(:,:,3);
figure(2),subplot(2,2,1),imshow(r,[]);title('r')
subplot(2,2,2),imshow(g,[]);title('g')
subplot(2,2,3),imshow(b,[]);title('b')
threshold=105;
BW = roicolor(b,threshold,255);
BW=imcomplement(BW);
figure(3);imshow(BW);
[L,num]=bwlabel(BW,4);
se=strel('disk',8,8);
se2=strel('disk',6,4);
BW=imdilate(BW,se);
BW=imerode(BW,se2);
figure(4),imshow(BW,[]);title('Morphed Image')
[L,num]=bwlabel(BW,8);
BW = imfill(BW,'holes');
figure(4),imshow(BW,[]);title('Morphed Image')
lab=RGB2Lab(Im);
lab_L=lab(:,:,1);
lab_a=lab(:,:,2);
lab_b=lab(:,:,3);
STATS=regionprops(L,b,'Centroid');
hold on
for k = 1:numel(STATS)
    centroid_k = STATS(k).Centroid;
    plot(centroid_k(1), centroid_k(2), 'b.');
end
hold off
[L,num]=bwlabel(BW,4);
STATS=regionprops(L,(r),'Area','PixelList','PixelValue','Centroid');
STATS2=regionprops(L,(lab_L),'Area','PixelList','PixelValue','Centroid'
);
```

```
STATS3=regionprops(L,(lab_a),'Area','PixelList','PixelValue','Centroid'
);
STATS4=regionprops(L,(lab_b),'Area','PixelList','PixelValue','Centroid'
);
STATS5=regionprops(L,(b),'Area','PixelList','PixelValue','Centroid');
STATS6=regionprops(L,(g),'Area','PixelList','PixelValue','Centroid');
STATS7=regionprops(L,imadjust(r),'Area','PixelList','PixelValue','Centr
oid');
STATS8=regionprops(L,imadjust(lab_L),'Area','PixelList','PixelValue','C
entroid');
STATS9=regionprops(L,imadjust(lab_a),'Area','PixelList','PixelValue','C
entroid');

STATS10=regionprops(L,imadjust(lab_b),'Area','PixelList','PixelValue','
Centroid');
STATS11=regionprops(L,imadjust(b),'Area','PixelList','PixelValue','Cent
roid');
STATS12=regionprops(L,imadjust(g),'Area','PixelList','PixelValue','Cent
roid');
 BeanNum=0;
 totalBean=0;
 [M, N, K]=size(g);
for bean = 1:num
    totalBean = totalBean+1;
    Area=cat(1,STATS.Area);
    F(totalBean+1,1)={totalBean} ;
    F(totalBean+1,2)={STATS(bean).Area};
end
centroid=ones(num,1);
centroid=cat(1,STATS.Centroid);
%RGB, Lab -domain value extraction of each pixel of each seed
pixelvalue_r=cat(1,STATS.PixelValues);
pixelvalue_l=cat(1,STATS2.PixelValues);
pixelvalue_a=cat(1,STATS3.PixelValues);
pixelvalue_b=cat(1,STATS4.PixelValues);
pixelvalue_g=cat(1,STATS5.PixelValues);
pixelvalue_blue=cat(1,STATS6.PixelValues);
pixelvalue_r_ad=cat(1,STATS7.PixelValues);
pixelvalue_l_ad=cat(1,STATS8.PixelValues);
pixelvalue_a_ad=cat(1,STATS9.PixelValues);
pixelvalue_b_ad=cat(1,STATS10.PixelValues);
pixelvalue_g_ad=cat(1,STATS11.PixelValues);
pixelvalue_blue_ad=cat(1,STATS12.PixelValues);
n=1;p=1;
 if (size(pixelvalue_r,1)-size(STATS(num).PixelValues,1))< 1
       j=1;
         m=size(STATS(j).PixelValues,1);
        pixel_r(1:m,p)=pixelvalue_r(n:m+n-1,1);
        pixel_g(1:m,p)=pixelvalue_g(n:m+n-1,1);
        pixel_blue(1:m,p)=pixelvalue_blue(n:m+n-1,1);
        pixel_l(1:m,p)=pixelvalue_l(n:m+n-1,1);
        pixel_a(1:m,p)=pixelvalue_a(n:m+n-1,1);
        pixel_b(1:m,p)=pixelvalue_b(n:m+n-1,1);
        pixel_r_ad(1:m,p)=pixelvalue_r_ad(n:m+n-1,1);
        pixel_g_ad(1:m,p)=pixelvalue_g_ad(n:m+n-1,1);
        pixel_blue_ad(1:m,p)=pixelvalue_blue_ad(n:m+n-1,1);
```

```
        pixel_l_ad(1:m,p)=pixelvalue_l_ad(n:m+n-1,1);
        pixel_a_ad(1:m,p)=pixelvalue_a_ad(n:m+n-1,1);
        pixel_b_ad(1:m,p)=pixelvalue_b_ad(n:m+n-1,1);
  else
     for j=1:(size(pixelvalue_r,1)-size(STATS(num).PixelValues,1))
        m=size(STATS(j).PixelValues,1);
        pixel_r(1:m,p)=pixelvalue_r(n:m+n-1,1);
        pixel_g(1:m,p)=pixelvalue_g(n:m+n-1,1);
         pixel_blue(1:m,p)=pixelvalue_blue(n:m+n-1,1);
        pixel_l(1:m,p)=pixelvalue_l(n:m+n-1,1);
        pixel_a(1:m,p)=pixelvalue_a(n:m+n-1,1);
        pixel_b(1:m,p)=pixelvalue_b(n:m+n-1,1);
        pixel_r_ad(1:m,p)=pixelvalue_r_ad(n:m+n-1,1);
        pixel_g_ad(1:m,p)=pixelvalue_g_ad(n:m+n-1,1);
        pixel_blue_ad(1:m,p)=pixelvalue_blue_ad(n:m+n-1,1);
        pixel_l_ad(1:m,p)=pixelvalue_l_ad(n:m+n-1,1);
        pixel_a_ad(1:m,p)=pixelvalue_a_ad(n:m+n-1,1);
        pixel_b_ad(1:m,p)=pixelvalue_b_ad(n:m+n-1,1);
        n=m+n;
        if n>size(pixelvalue_r,1)
            break;
        end
        p=p+1;
    end
 end

%removing seeds with area less than 65
xlswrite('Canola data',F,'Sheet1');
data = xlsread('Canola data');
data(:,3:4)=centroid;
area=data(:,2);
mean_area=sum(area)/size(area,1);
sigma_area=std(area);
outlier=(abs(area-mean_area))>2*sigma_area;
data(:,5)=outlier;
idx=find((data(:,5)==1)& (data(:,2)<400));
    data(idx,:)=[];
    pixel_r(:,idx)=[];
    pixel_l(:,idx)=[];
    pixel_a(:,idx)=[];
    pixel_b(:,idx)=[];
    pixel_g(:,idx)=[];
    pixel_blue(:,idx)=[];
    pixel_r_ad(:,idx)=[];
    pixel_l_ad(:,idx)=[];
    pixel_a_ad(:,idx)=[];
    pixel_b_ad(:,idx)=[];
    pixel_g_ad(:,idx)=[];
    pixel_blue_ad(:,idx)=[];

%identifying+removing brown pixels and calculating mean intensity of
remaining
%pixels in r,Lab-domain
s_r=uint32(zeros(1,size(pixel_r,2)));
s_l=uint32(zeros(1,size(pixel_l,2)));
s_a=uint32(zeros(1,size(pixel_a,2)));
s_b=uint32(zeros(1,size(pixel_b,2)));
```

```
s_g=uint32(zeros(1,size(pixel_g,2)));
s_blue=uint32(zeros(1,size(pixel_blue,2)));
s_r_ad=uint32(zeros(1,size(pixel_r,2)));
s_l_ad=uint32(zeros(1,size(pixel_l,2)));
s_a_ad=uint32(zeros(1,size(pixel_a,2)));
s_b_ad=uint32(zeros(1,size(pixel_b,2)));
s_g_ad=uint32(zeros(1,size(pixel_g,2)));
s_blue_ad=uint32(zeros(1,size(pixel_blue,2)));
n=(zeros(1,size(pixel_r,2)));
for i=1:1:size(pixel_r,1)
    for j=1:1:size(pixel_r,2)
        if pixel_r(i,j)>130
            true_pixel_r(i,j)=pixel_r(i,j);
             true_pixel_l(i,j)=pixel_l(i,j);
              true_pixel_a(i,j)=pixel_a(i,j);
               true_pixel_b(i,j)=pixel_b(i,j);
               true_pixel_g(i,j)=pixel_g(i,j);
               true_pixel_blue(i,j)=pixel_blue(i,j);
               true_pixel_r_ad(i,j)=pixel_r_ad(i,j);
             true_pixel_l_ad(i,j)=pixel_l_ad(i,j);
              true_pixel_a_ad(i,j)=pixel_a_ad(i,j);
               true_pixel_b_ad(i,j)=pixel_b_ad(i,j);
               true_pixel_g_ad(i,j)=pixel_g_ad(i,j);
               true_pixel_blue_ad(i,j)=pixel_blue_ad(i,j);
            n(1,j)=n(1,j)+1;
        end
    end
end

s_r=sum(true_pixel_r);
s_l=sum(true_pixel_l);
s_a=(sum(true_pixel_a));
s_b=sum(true_pixel_b);
s_g=sum(true_pixel_g);
s_blue=sum(true_pixel_blue);
s_r_ad=sum(true_pixel_r_ad);
s_l_ad=sum(true_pixel_l_ad);
s_a_ad=(sum(true_pixel_a_ad));
s_b_ad=sum(true_pixel_b_ad);
s_g_ad=sum(true_pixel_g_ad);
s_blue_ad=sum(true_pixel_blue_ad);
for j=1:1:size(s_r,2)
    mean_r(1,j)=(s_r(1,j)/n(1,j));
    mean_l(1,j)=(s_l(1,j)/n(1,j));
    mean_a(1,j)=((s_a(1,j))/n(1,j));
    mean_b(1,j)=(s_b(1,j)/n(1,j));
    mean_g(1,j)=(s_g(1,j)/n(1,j));
    mean_blue(1,j)=(s_blue(1,j)/n(1,j));
    mean_r_ad(1,j)=(s_r_ad(1,j)/n(1,j));
    mean_l_ad(1,j)=(s_l_ad(1,j)/n(1,j));
    mean_a_ad(1,j)=((s_a_ad(1,j))/n(1,j));
    mean_b_ad(1,j)=(s_b_ad(1,j)/n(1,j));
    mean_g_ad(1,j)=(s_g_ad(1,j)/n(1,j));
    mean_blue_ad(1,j)=(s_blue_ad(1,j)/n(1,j));
end
mean(:,1)= mean_r';
mean(:,2)= mean_g';
```

```
mean(:,3)= mean_blue';
mean(:,4)= mean_l';
mean(:,5)= mean_a';
mean(:,6)= mean_b';
mean(:,7)= mean_r_ad';
mean(:,8)= mean_g_ad';
mean(:,9)= mean_blue_ad';
mean(:,10)= mean_l_ad';
mean(:,11)= mean_a_ad';
mean(:,12)= mean_b_ad';
r=size(mean,1);
c=size(mean,2);
gd=zeros(r,c);
gr_d=zeros(r,c);
hd=zeros(r,c);
heat=0;
good=0;
green=0;
 for i=1:r
    if mean(i,5)<= -4.5
        green=green+1;
         gr_d(i,:)=mean(i,:);
    end
    if mean(i,5)<=-3.5 && mean(i,5)>= -4.5
        if(mean(i,1)>175 || mean(i,2)>91)
        good=good+1;
        gd(i,:)=mean(i,:);
    else
        green=green+1;
        gr_d(i,:)=mean(i,:);
        end
    end
    if mean(i,5)>= -3.5 && mean (i,5)<=0
        good=good+1;
         gd(i,:)=mean(i,:);
    end
     if mean(i,5)>=1.5
        heat=heat+1;
        hd(i,:)=mean(i,:);
    end
    if mean(i,5)>0 && mean(i,5)<1.5
        if mean(i,8)> 110
            good=good+1;
            gd(i,:)=mean(i,:);
        end
       if mean(i,8)<=110 && mean(i,7)<=150&& mean(i,2)<= 104 &&
mean(i,9)<= 130
                heat=heat+1;
                hd(i,:)=mean(i,:);
       else
           good=good+1;
            gd(i,:)=mean(i,:);
       end
         if  (mean(i,8)> 110 && mean(i,7)> 150)&& mean(i,2)<= 104
            heat=heat+1;
            hd(i,:)=mean(i,:);
```

```
        elseif ((mean(i,8)> 110&& mean(i,7)> 150)&& mean(i,2)> 104)&&
mean(i,9)<= 130
            heat=heat+1;
            hd(i,:)=mean(i,:);
             end
        end

end
```

APPENDIX II

FATTY ACID COMPOSITION OF SAMPLES

Sample	Palmitic	Stearic	Oleic	Linoleic	Linolenic	Erucic	Avg_Erucic
A-1.1	4.4453	1.6237	57.9024	15.9951	6.9578	0.2564	
A-1.2	5.2951	1.9450	66.4451	18.3943	7.9031	0.2691	
A-1.3	5.2849	1.9731	68.0643	18.8707	8.1747	0.2031	
A-2.1	3.9716	1.4368	48.9816	14.2511	6.1075	0.1539	
A-2.2	5.0853	1.7570	61.0318	16.7446	7.1339	0.2217	0.22
A-2.3	4.9856	1.8209	65.1583	18.1369	7.7717	0.2039	
A-3.1	-	-	-	-	-	-	
A-3.2	5.0728	1.9369	63.6518	17.6451	7.5662	0.2339	
A-3.3	4.6145	1.7310	62.2352	17.3170	7.4445	0.2382	
B-1.1	3.9342	1.3767	51.5328	14.4338	6.0422	0.6217	
B-1.2	4.7989	1.7146	63.0499	17.7558	7.4350	0.8503	
B-1.3	4.7664	1.7121	62.8619	17.5735	7.3716	0.7718	
B-2.1	5.1892	1.7464	63.2632	17.5948	7.3960	0.7619	
B-2.2	-	-	-	-	-	-	0.72
B-2.3	4.5267	1.6174	60.3868	16.8478	7.1275	0.7416	
B-3.1	4.1763	1.4736	55.6355	15.4978	6.5591	0.6488	
B-3.2	4.3701	1.5340	57.2205	16.0221	6.7699	0.6907	
B-3.3	10.1252	1.5632	57.9390	16.1440	6.8559	0.6654	
C-1.1	4.193372	1.654359	58.39873	15.1254	6.131130574	0	
C-1.2	4.335853	1.74143	61.88092	16.03039	6.540258239	0	
C-1.3	5.004268	1.945883	70.09911	18.11358	7.396088397	0	
C-2.1	4.213645	1.666455	59.60789	15.49477	6.280185964	0.234133745	
C-2.2	3.897719	1.583589	55.38924	14.4041	5.822088719	0	0.08
C-2.3	4.303413	1.711523	60.18877	15.72712	6.293411355	0	
C-3.1	4.02391	1.578851	55.05295	14.29525	5.764154891	0	
C-3.2							
C-3.3	4.413411	1.786817	62.54387	16.29706	6.827561133	0	

D-1.1	4.7716	1.6851	58.0694	15.9098	6.0922	0.1027	
D-1.2	5.0198	1.7566	60.4906	16.7221	6.4029	0.0765	
D-1.3	-	-	-	-	-	-	
D-2.1	4.6489	1.6705	56.7329	15.6539	6.0251	0.0789	0.08
D-2.2	4.6290	1.6570	56.9861	15.6674	5.9850	0.0000	
D-2.3	-	-	-	-	-	-	
D-3.1	4.8175	1.7366	59.1066	16.3111	6.2986	0.1063	
D-3.2	5.4429	1.9783	66.2377	18.2599	7.0266	0.1071	
D-3.3							
E-1.1	4.3260	1.6868	55.4255	15.6151	5.9951	0.0000	
E-1.2	1.5639	0.6645	18.6448	5.2710	2.0836	0.0000	
E-1.3	4.6009	1.5755	52.1841	14.8007	5.7084	0.0000	
E-2.1	5.4546	2.1486	69.4907	19.4199	7.5030	0.0000	
E-2.2	4.0693	1.6104	51.0914	14.3361	5.5608	0.1333	0.08
E-2.3	4.3217	1.8105	53.1521	14.8107	5.8335	0.2354	
E-3.1	50.6715	2.0417	64.1458	17.8673	6.8699	0.0955	
E-3.2	6.4327	2.5400	82.0333	22.8278	8.7552	0.0000	
E-3.3	4.0579	1.6666	51.0525	14.1979	5.5078	0.0000	
F-1.1	3.1306	1.3176	38.4028	10.7973	4.2004	0.0895	
F-1.2	5.0525	2.1369	64.4909	17.7354	6.7898	0.0000	
F-1.3	5.5444	1.6283	64.5502	18.0864	6.9497	0.0000	
F-2.1	5.1347	1.9937	64.2656	17.9766	6.7697	0.0000	
F-2.2	5.7267	2.2367	71.2675	19.9499	7.5369	0.0000	0.03
F-2.3	5.5576	2.2128	68.4766	19.0479	7.2481	0.0000	
F-3.1	5.7299	2.2093	72.9542	20.3871	7.7032	0.0413	
F-3.2	5.9846	2.3235	75.5076	21.1212	7.9951	0.0531	
F-3.3	5.4215	1.5886	64.7172	19.0113	7.2018	0.0229	
G-1.1	4.8717	2.2186	59.1328	18.0733	7.0517	0.0000	
G-1.2	5.2496	2.0740	68.6446	19.1020	7.1657	0.0000	
G-1.3	4.5201	1.5340	50.5961	14.2018	5.3473	0.0000	0.00
G-2.1	5.2122	2.0413	66.4693	18.3630	6.9134	0.0000	
G-2.2	5.2951	2.0718	67.7087	18.7371	7.0021	0.0000	
G-2.3	4.6803	1.6513	55.0400	16.0481	6.0046	0.0000	
H-1.1	5.0454	1.7899	63.3365	18.3094	8.4329	0.9336	
H-1.2	-	-	-	-	-	-	
H-1.3	5.0139	1.6056	56.4250	16.4327	7.6204	0.6933	
H-2.1	4.7996	1.6869	58.6205	16.9536	7.7517	0.9301	
H-2.2	5.3608	1.8479	63.2279	18.0565	8.1080	1.0034	1.02
H-2.3	-	-	-	-	-	-	
H-3.1	5.2567	1.8510	65.7782	19.0856	8.6175	1.3270	
H-3.2	4.9446	1.7074	60.8938	17.7612	8.1277	1.2377	
H-3.3	4.1356	1.3929	50.0863	14.4653	6.6492	1.0229	

2-1.1	4.0827	1.4538	44.9588	13.2825	5.4620	6.1614	5.94
2-1.2	5.7427	2.0504	63.6258	18.7673	7.7594	8.2706	
2-1.3	4.4402	1.3281	41.5223	12.3557	5.1435	4.7628	
2-2.1	3.9555	1.3912	43.1289	12.7636	5.3250	6.5364	
2-2.2	3.7768	1.2652	38.4233	11.2872	4.7156	5.9174	
2-2.3	3.9091	1.4209	43.5328	12.8393	5.2556	6.4394	
2-3.1	3.6747	1.3933	42.1040	12.3490	4.9872	5.4667	
2-3.2	4.0010	1.5254	46.6119	13.7047	5.5804	6.5216	
2-3.3	3.4026	1.3057	39.6036	11.5742	4.7419	5.7443	
3-1.1	4.0280	1.4618	37.1384	12.9349	6.0196	14.7644	10.75
3-1.2	-	-	-	-	-	-	
3-1.3	3.8503	1.2250	35.0974	11.5141	5.2805	9.0293	
3-2.1	3.8804	1.3985	37.8434	12.5906	5.5349	10.5123	
3-2.2	4.1437	1.4948	40.1714	13.3656	5.8773	11.2594	
3-2.3	3.8493	1.3524	39.4799	12.4492	5.4677	10.5097	
3-3.1	3.7001	1.3159	38.4543	12.1419	5.4504	11.1327	
3-3.2	3.9369	1.3990	40.8358	12.9143	5.8126	12.0300	
3-3.3	3.3641	1.1781	34.6757	10.9760	4.8963	10.2144	
4-1.1	3.0221	1.0567	25.4665	10.1965	5.5685	22.3234	20.86
4-1.2	-	-	-	-	-	-	
4-1.3	3.1994	0.9473	21.1719	9.2545	5.0631	15.9431	
4-2.1	3.0732	1.0565	25.8175	10.2512	5.5695	21.9273	
4-2.2	3.1098	1.0763	26.2094	10.4423	5.6753	22.4362	
4-2.3	-	-	-	-	-	-	
4-3.1	-	-	-	-	-	-	
4-3.2	3.5660	1.2402	30.1536	11.8188	6.3323	23.3912	
4-3.3	3.0162	1.0406	25.4406	10.0013	5.3485	19.1302	
5-1.1	3.1435	1.0407	20.5679	10.6510	6.7294	32.6370	33.48
5-1.2	3.5199	1.1579	23.3955	12.1297	7.6668	38.4310	
5-1.3	-	-	-	-	-	-	
5-2.1	3.2182	1.0458	109.0627	11.0783	7.0077	36.8003	
5-2.2	3.1009	1.0104	20.0969	10.5257	6.6343	34.2971	
5-2.3	3.0663	0.9996	19.7182	10.3248	6.4906	32.7184	
5-3.1	1.0533	0.3290	6.0796	2.9358	1.7894	9.0109	
5-3.2	2.3356	0.8060	14.9912	7.6935	4.8168	24.8666	
5-3.3	3.0864	1.0111	20.3141	10.5263	6.6313	34.6377	
6-1.1	2.1223	0.9561	9.7386	13.4844	12.0742	47.6249	46.11
6-1.2	1.9857	0.9712	8.7527	12.5238	11.2803	46.9490	
6-1.3	-	-	-	-	-	-	
6-2.1	1.8998	0.9573	8.8078	12.8788	11.7238	45.7458	
6-2.2	1.6953	0.8349	7.9089	12.4914	10.6909	45.1320	

6-2.3	1.8474	0.8707	8.4539	12.6418	11.4339	45.0748	
6-3.1	2.5104	1.1511	11.0440	16.6467	14.9740	59.5562	
6-3.2	-	-	-	-	-	-	
6-3.3	3.1592	1.5179	14.0360	21.0435	18.8608	79.7770	
7-1.1	2.9586	0.9433	12.8582	11.2057	8.4131	59.6814	52.63
7-1.2	2.8131	0.8747	12.2351	10.6098	7.9068	53.0219	
7-1.3	2.9472	0.8248	11.5648	10.0074	7.5152	40.2062	
7-2.1	3.2194	0.9874	13.5937	11.8263	8.8603	58.4167	
7-2.2	2.5165	0.7720	10.1041	9.3596	6.9779	45.4466	
7-2.3	2.9530	0.9015	11.8353	10.9209	8.1692	54.8536	
7-3.1	2.9823	0.9116	12.8883	11.0247	8.2645	53.8824	
7-3.2	2.6182	0.8272	10.8724	9.9290	7.4658	53.8879	
7-3.3	3.2230	0.9625	13.7499	11.6905	8.7699	54.2518	

APPENDIX III

FREE FATTY ACID CONTENT OF EACH SAMPLE

Sample #	Weight of Sample (gm)	Initial Volume (ml)	Final Volume (ml)	Vol. of NaOH used (ml)	%FFA
Sample 1					
1.1					
1.1.1	1.4107	0	1.9	1.9	0.95
1.1.2	1.4132	1.8	3.45	1.65	0.82
1.1.3	1.4117	3.6	5.1	1.5	0.75
1.2					
1.2.1	1.4147	0	1.6	1.6	0.80
1.2.2	1.413	1.6	3.2	1.6	0.80
1.2.3	1.413	3.1	4.55	1.45	0.72
1.3					
1.3.1	1.4131	4.6	6	1.4	0.70
1.3.2	1.413	6	7.45	1.45	0.72
1.3.3	1.4141	2	3.35	1.35	0.67
Sample 2					
2.1					
2.1.1	1.4103	0.5	4.45	3.95	1.97
2.1.2	1.4113	3	6.85	3.85	1.92
2.1.3	1.4113	1.5	5.05	3.55	1.77
2.2					
2.2.1	1.411	0.5	4.2	3.7	1.85
2.2.2	1.4139	4	7.7	3.7	1.84
2.2.3	1.4125	3	6.6	3.6	1.80
2.3					
2.3.1	1.4117	0	3.65	3.65	1.82
2.3.2	1.4111	3	6.65	3.65	1.82
2.3.3	1.4114	2	5.65	3.65	1.82

Sample 3					
3.1					
3.1.1	1.4109	0	2.5	2.5	1.25
3.1.2	1.4098	0.5	4.6	4.1	2.05
3.1.3	1.4109	1.5	4.5	3	1.50
3.2					
3.2.1	1.412	0	2.15	2.15	1.07
3.2.2	1.4113	2.15	4.35	2.2	1.10
3.2.3	1.4105	4.35	6.6	2.25	1.12
3.3					
3.3.1	1.4132	0	2.5	2.5	1.25
3.3.2	1.4101	2.55	4.95	2.4	1.20
3.3.3	1.4114	4.95	7.35	2.4	1.20
4.1					
4.1.1	1.4105	0	3.45	3.45	1.72
4.1.2	1.41	3.45	6.5	3.05	1.53
4.1.3	1.4105	6.3	9.35	3.05	1.52
4.2					
4.2.1	1.4102	0	3.2	3.2	1.60
4.2.2	1.4106	3.2	6.7	3.5	1.75
4.2.3	1.4104	6.7	10	3.3	1.65
4.3					
4.3.1	1.4107	0	3.25	3.25	1.62
4.3.2	1.4104	3.25	6.8	3.55	1.77
4.3.3	1.4101	4	7.5	3.5	1.75
Sample 5					
5.1					
5.1.1	1.41	0	3.2	3.2	1.60
5.1.2	1.412	3.2	5.85	2.65	1.32
5.1.3	1.4121	0	3.1	3.1	1.55
5.2					
5.2.1	1.4113	1.5	4.35	2.85	1.42
5.2.2	1.4108	4.35	7.2	2.85	1.42
5.2.3	1.4116	0	2.85	2.85	1.42
5.3					
5.3.1	1.411	0	2.85	2.85	1.42
5.3.2	1.41	2.85	5.85	3	1.50
5.3.3	1.4114	5.85	8.8	2.95	1.47
Sample6					
6.1					
6.1.1	1.4106	0	0.35	0.35	1.75
6.1.2	1.4101	0.35	0.6	0.25	1.25
6.1.3	1.4107	0	3.05	3.05	1.52
6.2					
6.2.1	1.4099	1	3.8	2.8	1.40
6.2.2	1.4107	3.8	6.65	2.85	1.42
6.2.3	1.4106	0	2.7	2.7	1.35

6.3					
6.3.1	1.4099	5.35	8.25	2.9	1.45
6.3.2	1.4101	1	1.3	0.3	1.50
6.3.3	1.4108	2.5	5.35	2.85	1.42
Sample 7					
7.1					
7.1.1	1.4099	0	3.3	3.3	1.65
7.1.2	1.4105	3.3	6.55	3.25	1.62
7.1.3	1.4106	0	3.25	3.25	1.62
7.2					
7.2.1	1.4102	3.25	6.55	3.3	1.65
7.2.2	1.4101	0.5	3.65	3.15	1.57
7.2.3	1.4105	2	5.35	3.35	1.67
Sample 8					
8.1					
8.1.1	1.4101	0	2.5	2.5	1.25
8.1.2	1.41	2.5	5.15	2.65	1.33
8.1.3	1.41	0	2.5	2.5	1.25
8.2					
8.2.1	1.41	2.5	5.2	2.7	1.35
8.2.2	1.41	5.2	7.9	2.7	1.35
8.2.3	1.41	0	2.45	2.45	1.23
8.3					
8.3.1	1.41	2	4.5	2.5	1.25
8.3.2	1.41	4.7	7.15	2.45	1.23
8.3.3	1.41	0	2.55	2.55	1.28

APPENDIX IV

PEROXIDE VALUE OF EACH SAMPLE

Sample #	Weight of sample (gm)	Initial volume (ml)	Final Volume (ml)	Blank	Vol. of $Na_2S_2O_3$ used (ml)	PV
Sample 1						
1.1						
1.1.1	0.5096	0	2.95	0	2.95	5.79
1.1.2	0.5024	1	3.95	0	2.95	5.87
1.1.3	0.5006	4	6.95	0	2.95	5.89
1.2						
1.2.1	0.5025	7	9.4	0	2.4	4.78
1.2.2	0.5028	1.5	4.25	0	2.75	5.47
1.2.3	0.502	4.25	6.3	0	2.05	4.08
1.3						
1.3.1	0.5002	3	3.95	0	0.95	1.90
1.3.2	0.4995	3.95	5.1	0	1.15	2.30
1.3.3	0.5023	5.1	6.15	0	1.05	2.09
Sample 2						
2.1						
2.1.1	0.5028	0	3.7	0	3.7	7.36
2.1.2	0.5018	3.5	7.3	0	3.8	7.57
2.1.3	0.5002	1	4.8	0	3.8	7.60
2.2						
2.2.1	0.5033	2.75	5.5	0	2.75	5.46
2.2.2	0.5017	5.5	8.3	0	2.8	5.58
2.2.3	0.5007	0	2.75	0	2.75	5.49
2.3						
2.3.1	0.5007	0.5	3.05	0	2.55	5.09
2.3.2	0.5005	2.9	5.5	0	2.6	5.19
2.3.3	0.5009	5.65	8.8	0	3.15	6.29

Sample 3						
3.2						
3.2.1	0.5003	0	4	0	4	8.00
3.2.2	0.5009	0	3.4	0	3.4	6.79
3.3						
3.3.1	0.5007	0	4.1	0	4.1	8.19
3.3.2	0.5003	0	4	0	4	8.00
3.3.3	0.5002	0	3.95	0	3.95	7.90
Sample 4						
4.1						
4.1.1	0.4998	0	0.9	0	0.9	1.80
4.1.2	0.5008	0.9	1.8	0	0.9	1.80
4.1.3	0.5001	1.8	2.7	0	0.9	1.80
4.2						
4.2.1	0.5008	0	0.7	0	0.7	1.40
4.2.2	0.4998	1	1.75	0	0.75	1.50
4.2.3	0.5001	2	2.65	0	0.65	1.30
4.3						
4.3.1	0.5001	3	3.65	0	0.65	1.30
4.3.2	0.5002	3.65	4.2	0	0.55	1.10
4.3.3	0.5002	5	5.7	0	0.7	1.40
Sample 5						
5.1						
5.1.1	0.4999	0	1.95	0	1.95	3.90
5.1.2	0.5002	1.95	3.55	0	1.6	3.20
5.1.3	0.5004	3.55	5.2	0	1.65	3.30
5.2						
5.2.1	0.5007	0	0.85	0	0.85	1.70
5.2.2	0.4999	1	1.9	0	0.9	1.80
5.2.3	0.5005	1.9	2.6	0	0.7	1.40
5.3						
5.3.1	0.4998	2.6	3.25	0	0.65	1.30
5.3.2	0.5005	3.5	4	0	0.5	1.00
5.3.3	0.5005	4	4.7	0	0.7	1.40
Sample6						
6.1						
6.1.1	0.5004	0	2.75	0	2.75	5.50
6.1.2	0.4999	2.75	5.7	0	2.95	5.90
6.1.3	0.4998	5.7	8.45	0	2.75	5.50
6.2						
6.2.1	0.5002	0	1	0	1	2.00
6.2.2	0.5002	1	3.6	0	2.6	5.20
6.2.3	0.5001	3.6	6.25	0	2.65	5.30
6.3						
6.3.1	0.5007	6.25	8.95	0	2.7	5.39
6.3.2	0.5007	1	3.65	0	2.65	5.29
6.3.3	0.5001	2	4.6	0	2.6	5.20

Sample 7						
7.1						
7.1.1	0.5001	0	1.8	0	1.8	3.60
7.1.2	0.5005	1.8	3.6	0	1.8	3.60
7.1.3	0.5001	3.6	5.25	0	1.65	3.30
7.2						
7.2.1	0.5003	5.25	7.2	0	1.95	3.90
7.2.2	0.5004	0	1.9	0	1.9	3.80
7.2.3	0.5006	1.9	4	0	2.1	4.19
Sample 8						
8.1						
8.1.1	0.4999	1.5	4.25	0	2.75	5.50
8.1.2	0.5002	1	3.8	0	2.8	5.60
8.1.3	0.5005	3.8	6.75	0	2.95	5.89
8.2						
8.2.1	0.5	0	2.55	0	2.55	5.10
8.2.2	0.5001	2.55	5.45	0	2.9	5.80
8.2.3	0.5005	5.45	8.35	0	2.9	5.79
8.3						
8.3.1	0.4998	0	2.5	0	2.5	5.00
8.3.2	0.5008	2.5	5.25	0	2.75	5.49
8.3.3	0.5005	5.25	8.1	0	2.85	5.69

VITA

Geetika Dilawari

Candidate for the Degree of

Doctor of Philosophy/Biosystems Engineering

Thesis: QUALITY ESTIMATION OF CANOLA USING MACHINE VISION AND VIS-NIR SPECTROSCOPY

Major Field: Biosystems and Agricultural Engineering

Education:

Completed the requirements for the Doctor of Philosophy in Biosystems and Agricultural Engineering at Oklahoma State University, Stillwater, Oklahoma in December, 2011.

Completed the requirements for the Master of Science/Arts in Agricultural and Biosystems Engineering at Iowa State University, Ames, Iowa in 2006.

Completed the requirements for the Bachelor of Science in Agricultural Engineering at Punjab Agricultural Univeristy, Ludhiana, Punjab, India in 2004.

Experience:

Graduate Research Assistant June 2010- Dec 2011
Oklahoma State University, Department of Biosystems and Agricultural Engineering

Research Engineer (Sensors and Control) Jan 2007-May 2010
Oklahoma State University, Department of Biosystems and Agricultural Engineering

Graduate Research Assistant Aug 2004-Dec 2006
Iowa State University, Department of Agricultural and Biosystems Engineering

Professional Memberships: ASABE, 2005-2010

Name: Geetika Dilawari Date of Degree: December, 2011

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: QUALITY ESTIMATION OF CANOLA USING MACHINE VISION AND VIS-NIR SPECTROSCOPY

Pages in Study: 91 Candidate for the Degree of Doctor of Philosophy/ Biosystems Engineering

Major Field: Biosystems and Agricultural Engineering

Scope and Method of Study:

Canola is mainly graded either by visual inspection or by smelling. These methods are subjective in nature and are bound to cause errors while deciding the grade of canola. To test canola for amount of erucic acid present the sample needs to be sent to a laboratory for testing through wet chemical analysis. This is a time consuming process. An electronic method that can quantify amount of dockage, presence of distinctly green and heat treated seeds, distinguish samples on the basis of erucic acid, its free fatty acid content and PV, would not only be less time consuming but also would be a more reliable method to grade canola samples.

Findings and Conclusions:

1. Canola samples cannot be classified on the basis of total dockage present using L and RGB data obtained from flat-bed scanner. Inclusion of morphological and textural features would improve the classification accuracy
2. Machine vision can be considered as a potential method to grade canola on the basis of good, distinctly green and heat damaged seeds.
3. NIR spectroscopy can be used to classify samples on the basis of erucic acid.
4. FFA and PV can be estimated using PLS model applied on NIR spectral data. Potential wavelengths responding to change in FFA and PV were also identified. These wavelengths can be used as indicators for lipid oxidation (rancidity) of canola oil.

ADVISER'S APPROVAL: Dr Carol Jones