DETERMINATION OF MAIN CONSTITUENTS IN WHEAT USING NEAR INFRARED HYPERSPECTRAL IMAGING

BY

MAHESH SIVAKUMAR

A Thesis Submitted to
The Faculty of Graduate Studies
In Partial Fulfillment of the Requirements for the Degree of

Master of Science

Department of Biosystems Engineering
University of Manitoba
Winnipeg, Manitoba

The Faculty of Graduate Studies
500 University Centre, University of Manitoba
Winnipeg, Manitoba R3T 2N2

Phone: (204) 474-9377
Fax: (204) 474-7553
graduate_studies@umanitoba.ca

THE UNIVERSITY OF MANITOBA

FACULTY OF GRADUATE STUDIES

*****

COPYRIGHT PERMISSION

DETERMINATION OF MAIN CONSTITUENTS IN WHEAT

USING NEAR INFRARED HYPERSPECTRAL IMAGING

BY

MAHESH SIVAKUMAR

A Thesis/Practicum submitted to the Faculty of Graduate Studies of The University of

Manitoba in partial fulfillment of the requirement of the degree

MASTER OF SCIENCE

ABSTRACT

Differentiation of wheat classes and rapid measurement of main constituents (e.g., protein, starch, oil content, and moisture content) in wheat are important challenges to the grain industry. In this study, NIR reflectance and absorbance values of hyperspectral images of wheat samples were used for identifying the Canadian wheat classes at same and different moisture levels and for predicting protein and oil content of wheat. Images of wheat were obtained using a NIR hyperspectral imaging system. Seventy five normalized NIR mean reflectance and NIR absorbance features were extracted from the scanned images of wheat. The extracted features were used to develop classification models and prediction models for identifying wheat classes; and predicting protein and oil contents of wheat, respectively.

Classification accuracies were 100% in classifying Canada Prairie Spring Red (CPSR), Canada Western Extra Strong (CWES), Canada Western Hard White Spring (CWHWS), Canada Western Red Spring (CWRS), Canada Western Red Winter (CWRW), and Canada Western Soft White Spring (CWSWS) wheat; and > 98% for the other two wheat classes (Canada Prairie Spring White (CPSW) and Canada Western Amber Durum (CWAD)) using linear discriminant analysis (LDA) with leave-one-out cross validation. Using quadratic discriminant analysis (QDA) with leave-one-out cross validation, the classification accuracies were > 96% for all wheat classes. The classification accuracies of 80 – 100% and 89 – 100% were found for artificial neural network (ANN) models with two different training patterns such as 60% training – 30% test – 10% validation (60-30-10) and 70% training – 20% test – 10% validation (70-20-10), respectively.

Classification accuracies of 100% were achieved using LDA with leave-one-out cross validation for CWSWS wheat at 14, 16, 18, and 20% moisture levels with 75 features. And, classification accuracies of > 90% were achieved for all wheat classes except CWES wheat at 20% moisture level and CWHWS wheat at 14% moisture level in LDA with leave-one-out cross validation using 75 features. Plots of the first two canonical variables showed that protein and moisture contents of wheat could be predicted using the NIR absorbance values of hyperspectral images. Principal components analysis (PCA) and STEPDISC procedure were used to find the top wavelengths in wheat class identification.

A 75 feature PLSR model for predicting protein in wheat produced the best standard error of prediction (SEP = 0.68) and a good correlation (r = 0.94) with the measured protein in wheat. Also, the 75 feature PLSR model for predicting oil content in wheat produced the best SEP of 0.10 and a r value of 0.83 with the measured oil content in wheat. Results of this study showed that NIR hyperspectral imaging could be used as an effective method for predicting protein and oil contents in wheat and identifying wheat classes at different moisture levels.

ACKNOWLEDGMENTS

It is my privilege to express my sincere gratitude to Dr. Digvir S. Jayas (Associate Vice-President (Research), University of Manitoba), my advisor, for his valuable guidance, suggestions, and support throughout the period of my research. He gave me an opportunity to continue my graduate studies at the University of Manitoba, Winnipeg. He also gave me a chance to work in NIR hyperspectral imaging in wheat quality determination as a part of my graduate program.

I also express my heartfelt thanks to Dr. Jitendra Paliwal (Assistant Professor, Department of Biosystems Engineering, University of Manitoba), my co-advisor, for his valuable guidance to complete this research successfully. He provided me with a lot of technical suggestions and ideas throughout the period of my research.

I thank Dr. Noel D.G. White (Senior Research Scientist, Agriculture and Agri-Food Canada, Winnipeg) for providing me lots of technical advice and helping me to get all necessary things for this research in time to complete it successfully.

I thank Dr. Reza Fazel (Assistant Professor, Department of Electrical and Computer Engineering, University of Manitoba) for serving on my advisory committee and reviewing my research work. I would like to say my special thanks to Dr. Manickavasagan Annamalai (Scientist, McCain Foods) for his help and sharing his knowledge for the past two years.

I would like to acknowledge the help I received from the other faculty members of the Department of Biosystems Engineering, University of Manitoba. I would like to thank the fellow research colleagues Chelladurai Vellaichamy, Chandra Bhan Singh, Dr. Prabal K. Ghosh, Dr. Ruplal Choudhary, Dr. Hao Zhang, Feng Wang, Gayathri Pitchai,

Hai Yan Li, Ke Sun, Nithya Udhayakumar, Rajarammanna Ramachandran, Sathya Gunasekaran, Sujala Balaji, Suresh Raja Neethirajan, Vadivambal Rajagopal, and Wu Jun for extending their support for the past two years. I sincerely thank all technical and administrative staff, Department of Biosystems Engineering for their help in this research.

Financial assistance from the Natural Sciences and Engineering Research Council of Canada is gratefully acknowledged.

I sincerely thank my parents, my brother, and my wife for providing me all kind of supports throughout the period of my research. At last, I would like to extend my thanks to all who provided support directly or indirectly for this research.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

LIST OF ACRONYMS

AACC	American Association of Cereal Chemists
ANN	Artificial Neural Network
AOAC	Association of Official Analytical Chemists
ASAE	American Society of Agricultural Engineers
ASCC	Average Squared Canonical Correlation
BAR	Board of Appeals and Review
BPNN	Back Propagation Neural Network
CPSR	Canada Prairie Spring Red
CPSW	Canada Prairie Spring White
CWAD	Canada Western Amber Durum
CWES	Canada Western Extra Strong
CWHWS	Canada Western Hard White Spring
CWRS	Canada Western Red Spring
CWRW	Canada Western Red Winter
CWSWS	Canada Western Soft White Spring
DHV	Dark Hard Vitreous
DON	Deoxynivalenol
GA	Genetic Algorithm
GC	Gas Chromatography
GUI	Graphic User Interface
HPLC	High Performance Liquid Chromatography
HRS	Hard Red Spring
HRW	Hard Red Winter
HWH	Hard White
LCTF	Liquid Crystal Tunable Filter
LDA	Linear Discriminant Analysis
MAE	Mean Absolute Error
Mid-IR	Mid infrared
MLR	Multiple Linear Regression
MNF	Minimum Noise Fraction
MSC	Multiplicative Scatter Correction
MSE	Mean Squared Error
NDHV	Non Dark Hard Vitreous
NIR	Near-infrared
PAGE	Polyacrylamide Gel Electrophoresis
PC	Principal Components
PCA	Principal Components Analysis
PCR	Principal Components Regression
PLS	Partial Least Squares
PLSR	Partial Least Squares Regression
PRESS	Predicted Error Sum of Squares
QDA	Quadratic Discriminant Analysis
RER	Ratio of standard error of performance to data range
RMSEC	Root Mean Square Error for Calibration

RMSECV	Root Mean Square Error for Cross Validation
RMSEP	Root Mean Square Error for Prediction
RPD	Relative Performance Determinant or Ratio of performance to deviation
RP-HPLC	Reversed Phase-High Performance Liquid Chromatography
SAS	Statistical Analysis Software
SD	Standard Deviation
SDS	Sodium Dodecyl Sulfate
SEP	Standard Error of Prediction
SIMCA	Soft Independent Modeling of Class Analogy
SKCS	Single Kernel Characterization System
SME	Specific Mechanical Energy
SMV	Soybean Mosaic Virus
SNV-DET	Standard Normal Variate and Detrending
SRW	Soft Red Winter
SSC	Soluble Solids Content
SWH	Soft White
US-FDA	United States-Food and Drug Administration
WILMA	Wavelet Interface to Linear Modeling Analysis

1. INTRODUCTION

Wheat is one of the main sources of food around the world. Wheat production and export in Canada were 25.9 Mt and 15.1 Mt respectively, in 2004 (FAOSTAT 2006). Production of several wheat classes while maintaining a zero tolerance policy in insect levels during export satisfy different types of consumers and produce maximum returns to the farmers in Canada. Specific wheat classes are used as primary ingredients for specific products like bread, pasta, flat breads, or noodles. Each wheat class has its own physical and chemical properties. Chemical compositions of wheat may be different for wheat classes that are similar in appearance. Accidental mixing of wheat classes during transportation or handling reduces the value of the lot considerably. A major challenge to the grain industry is accurate identification and differentiation of wheat classes at various moisture levels; and rapid measurement of the main constituents in wheat. Once wheat classes are identified, quality analyses are done on the sample to determine the constituents such as protein, starch, oil, and moisture content.

Wheat classes are broadly classified by their color (white or red), hardness (soft or hard), and growing season (spring or winter). Eight important wheat classes grown in western Canada are Canada Western Red Spring (CWRS), Canada Prairie Spring Red (CPSR), Canada Western Extra Strong (CWES), Canada Western Red Winter (CWRW), Canada Prairie Spring White (CPSW), Canada Western Amber Durum (CWAD), Canada Western Soft White Spring (CWSWS), and Canada Western Hard White Spring (CWHWS) (CGC 2006a).

Various methods have been used to identify and differentiate wheat classes. In grain handling facilities, a visual method is used to identify and differentiate the wheat

classes. Wheat classes and varieties can be identified by measuring specific proteins using polyacrylamide gel electrophoresis (PAGE) and high performance liquid chromatography (HPLC) (CGC 2007a). In PAGE, the sample proteins are either positively or negatively charged based on the pH of the medium and electrophoresed at specific voltage (Rybicki and Purves 2007). The separated protein bands are further subjected to staining and destaining processes and visualized under white illuminator (ICC 2001). Sodium dodecyl sulfate PAGE (SDS-PAGE) helps to separate proteins depending on their molecular weight and specific charge (Anonymous 2007a). SDS acts as a denaturing agent, thereby changing the shape of the protein molecules by splitting the hydrophobic linkages and imparting negative charge. A machine vision approach has been used to distinguish two wheat classes (CWRS and CWAD), barley, oats, and rye (Paliwal et al. 1999). Hard red spring (HRS) and hard red winter (HRW) wheat classes have been differentiated by quantitative estimation of gliadin, one of the four major protein components of wheat (Steinman 2007), using reversed-phase HPLC (RP-HPLC) (Huebner et al. 1995).

The grading and classification system of wheat in Canada is well established and ensures the primary grade determinants such as contrasting classes, diseased kernels, etc. to a permissible level (Fowler 2007). This system helps to preserve the commercial value of wheat by maintaining its processing potential and assuring maximum return to the producers (Dexter and Marchylo 2007). The development of a rapid method is necessary to identify wheat classes that would benefit producers, grain handlers, wheat millers, and processors (Dexter and Marchylo 2007). Canadian Grain Commission (CGC) wheat grades assure satisfactory performance in quality and milling potential and are considered

as the best wheat grades around the world (CWB 2007). Inter class identification and intra class grading are the two important activities in wheat marketing (Williams 2006).

Kjeldahl and combustion nitrogen methods are commonly used to determine the total nitrogen content in wheat. Conversion factors are used to find the total protein content of various agricultural crops from the obtained total nitrogen values. Conversion factors are 5.83, 6.25, 5.83, and 5.95 for some important crops such as barley, corn, millets, and rice, respectively. Whole wheat, germ, and endosperm of wheat have specific conversion factors such as 5.83, 6.31, and 5.70, respectively (FAO 2007).

Solvent extraction and gravimetric methods are used for fat content determination in agricultural commodities (FAO 2007). Crude fat and dietary fats are determined using a solvent extraction method with petroleum ether as solvent and a gravimetric method, respectively. Wet protein and fat analyses are complex and time consuming procedures that utilize hazardous chemicals and result in the sample being destroyed. There is a need for an effective alternate method to determine the main constituents in wheat without the drawbacks of traditional techniques.

NIR spectroscopy is used in various fields such as animal husbandry, agriculture, and pharmaceuticals. In the agriculture sector, it helps in determining quality parameters such as protein, moisture content, and oil content of whole (Delwiche 1998) and ground wheat (Wang et al. 2004). NIR spectroscopy has been used to determine deoxynivalenol (DON) mycotoxin levels in wheat (Petterson and Aberg 2003) and barley (Ruan et al. 2002) and identify waxy wheat varieties (Delwiche and Graybosch 2002). This technique has also been used to determine four different life stages of *Sitophilus oryzae* (L.) (rice weevil) at four different levels in artificially infested CWRS wheat (Paliwal et al. 2004).

A PLS model was developed using the data obtained from single kernel NIR equipment to determine both protein and moisture contents in soybean and only moisture content in corn (Armstrong 2006).

NIR hyperspectral imaging is the combination of two important techniques such as machine vision and NIR spectroscopy. NIR hyperspectral imaging data contain both spatial and spectral information in the form of a hypercube with two spatial dimensions and one spectral dimension (Lu and Chen 1998). This technique has the potential to develop effective models to determine various quality parameters of agricultural commodities. Good calibration results for measuring moisture and oil contents were obtained from single kernel analysis of maize using NIR hyperspectral imaging (Cogdill et al. 2004). This technique has been used for measuring the firmness and soluble solids content in strawberries (Nagata et al. 2005), detection of bruises, contaminations, and defects in apples (Lu 2003; Mehl et al. 2004), and identification of bitter pit lesions in apples (Nicolai et al. 2006).

The objectives of this study were:

1. to differentiate western Canadian wheat classes at same moisture level using statistical and ANN classifiers with normalized NIR mean reflectance intensities as input features,
2. to differentiate western Canadian wheat classes with and without moisture effects using statistical and ANN classifiers using NIR absorbance values as input features, and
3. to develop PLSR models to predict main constituents of wheat such as protein and oil content using NIR absorbance values as input features.

2. REVIEW OF LITERATURE

Consumers always prefer good quality products. Production of high quality products lead to marketing consistency and enhanced food safety. There are many techniques available to ensure online quality monitoring of an agricultural product. Standard chemical methods and various imaging techniques help to assess the quality and safety of agricultural products. The value of end products depends on the type of raw material, processing and storage methods used. Currently, imaging techniques are used for online grading in certain sectors and provide an alternate method to traditional quality measurement techniques. Some of the important imaging techniques are NIR or Vis/NIR hyperspectral imaging, monochrome imaging, color imaging, soft X-ray imaging, magnetic resonance imaging, X-ray computed tomography imaging, and thermal imaging.

2.1 Advantages and drawbacks of traditional methods

The traditional methods of measuring quality of agricultural commodities have some important advantages and drawbacks. Traditional methods are mostly direct methods and require unskilled or partially skilled labors, minimum statistical knowledge, handling of small amount of data, or minimum initial investment. Traditional methods are time consuming and subjective and require hazardous chemicals, high operating cost and destructive sample preparation. NIR hyperspectral imaging could be used as an effective alternate method to overcome some of the drawbacks of traditional methods.

2.2 Advantages and drawbacks of NIR hyperspectral imaging

NIR hyperspectral imaging is a type of NIR imaging in which images of a sample are taken at different wavelengths of the NIR region. NIR hyperspectral imaging shares similar advantages of NIR spectroscopy. This technique requires minimum time for quality measurement, minimum operating cost, zero chemicals, and non destructive sample preparation, and results more objective, comparatively accurate, and consistent. Requirement of high initial investment, proficient image and spectral data processing knowledge, and expertise in statistical tools handling are some of the important drawbacks of this technique.

2.3 NIR spectroscopy in the food industry

NIR spectroscopy is a modern technique and used for determining several quality parameters in raw and processed grains, milk, meat, fish, fruits, vegetables, confectionery, and beverages. It is also used for finding out the adulteration and genuineness of food products. Functional, compositional, and sensory analyses of food products could be done using NIR spectroscopy. The commonly used instruments for this technique are monochromators (H1034B, Jobin Yvon Inc., Edison, NJ), diode array spectrometers (8452A, Hewlett-Packard Inc., Palo Alto, CA), and filter instruments (MIR 06, Cambridge Research and Instrumentation Inc., Woburn, MA). Diffuse reflectance and diffuse transmittance are the two important methods used in this technique to find the NIR absorbance values of samples. Food absorbs NIR radiation based on the presence of main constituents. Molecular vibrations are excited in the sample in the form of combination bands and overtones of carbon, nitrogen, and hydrogen molecules based on the amount of NIR absorption. NIR absorbance, reflectance, transmittance, or a

combination of these data are pre-processed using standard normal variate and detrending (SNV-DET) and multiplicative scatter correction (MSC) and processed using different model calibration techniques such as multiple linear regression (MLR), principal components regression (PCR), and partial least squares regression (PLSR) (Osborne 2006). NIR absorption bands are shown in Fig. 2.1.

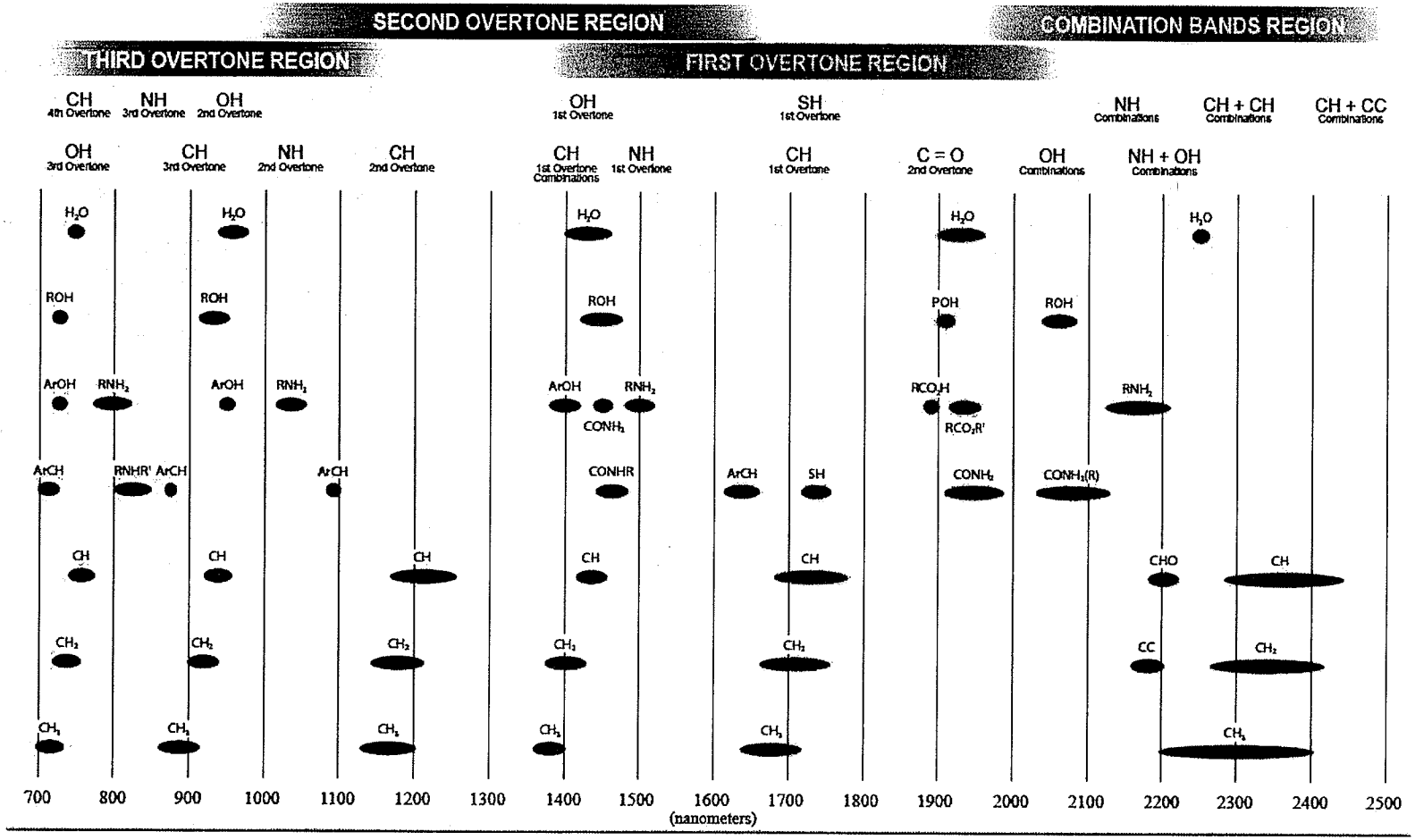

Fig. 2.1 NIR absorption bands.
Source: Anonymous (2007b).

2.3.1 Physical quality assessment of agricultural commodities using NIR spectroscopy

Hareland (1994) developed a NIR prediction model for determining the percent volume of flour particles in wheat flour derived from different wheat types and milling methods. NIR reflectance spectroscopy was used in this study and percent volume of flour particles was determined using a laser diffraction method. An NIR prediction model was developed using a PLSR technique and 96% of accuracy was reported for the wheat flours in determining the percent volume of flour particles.

Identification of waxy wheat and differentiation of waxy wheat from partially waxy and wild wheat varieties were studied using NIR spectroscopy (Delwiche and Graybosch 2002). NIR reflectance spectra of ground wheat samples were collected using a spectrophotometer in the NIR wavelength region of 1100 – 2498 nm at 2 nm wavelength intervals. Iodine binding blue complex colorimetric method was used to measure the apparent amylose content in wheat samples and the measured data were used as a reference data set. PCA was used for developing a wheat classification model. Spectral data were reduced to 15 principal components and stepwise regression was performed for the gene class separation. One-out cross validation was used for determining the optimal number of discriminant functions. Perfect classification could not be achieved using an NIR model because amylase contents were laid over the other gene classes in wheat.

Delwiche and Massie (1996) used Vis/NIR reflectance techniques to classify single kernels of wheat. Wheat samples from three hard wheat varieties (hard white (HWH), hard red spring (HRS), and hard red winter (HRW)) and two soft wheat varieties

(soft red winter (SRW) and soft white (SWH)) in U.S. were used for this study. A diode array spectrometer was used for collecting the reflectance intensities at the Vis/NIR wavelength region. NIR reflectance spectrophotometer was used to collect the reflectance intensities from the NIR wavelength region of 1100 – 2498 nm. PLSR and MLR methods were used for model development. Model classification accuracy was > 97% in differentiating red and white wheat classes using a seven factor PLSR model. For a five wavelength MLR model, classification accuracy of > 96% was reported for differentiating red and white wheat classes.

Cocchi et al. (2006) studied the feasibility of NIR spectroscopy to measure the degree of adulteration in durum wheat flour which was adulterated with common bread wheat flour. PLSR and wavelet interface to linear modeling analysis (WILMA) were used for developing models using raw data and data pretreated using SNV as input. A spectrophotometer was used for acquiring the spectra in the wavelength region of 400 – 2498 nm at 2 nm intervals. Spectral pretreatment helped to reduce the root mean square error for calibration (RMSEC) (= 0.2903), root mean square error of cross validation (RMSECV) (= 0.7215), and root mean square error of prediction (RMSEP) (= 0.3974) values of PLSR model developed with 8 latent variables. WILMA-PLS model with 60 coefficients and 7 latent variables had the minimum RMSEP value of 0.447 in quantifying the degree of adulteration in wheat flour.

Dowell (2000) used NIR spectroscopy to differentiate the vitreous and non vitreous kernels of durum wheat. A diode array spectrometer was used to collect the absorbance values from samples in the wavelength region of 400 – 1700 nm at 5 nm intervals. Two groups of wheat kernels (easily distinguishable group and easily non

distinguishable group) were formed and a two class PLSR model was developed for differentiating the wheat kernels based on vitreousness. Improved performance in identifying vitreous wheat kernels was achieved by increasing the discriminant value from 0.5 to 0.7 in the PLSR model. Use of a subjective reference method (Bureau of Appeals and Review method) in identifying vitreousness of wheat was the main cause for the performance reduction in the PLSR model. When a discriminant value of 0.5 was used, approximately 80% of vitreous and 70% of non vitreous kernels were classified correctly. NIR spectroscopy could be used in quantifying the durum wheat vitreousness because of the difference in NIR absorption by protein and starch contents of wheat.

Wang et al. (2002) studied the feasibility of Vis/NIR spectroscopy to detect dark hard vitreous (DHV) wheat kernels from non dark hard vitreous wheat (NDHV) kernels. Diode array spectrometer was used to collect the reflectance intensities from wheat kernel samples in the wavelength region of 400 – 1700 nm. This wavelength region was segmented into three portions such as 500 – 750 nm (visible region), 750 – 1700 nm (NIR region), and 500 – 1700 nm (Vis/NIR region) for calibration purposes. A two class PLSR model was developed for DHV wheat detection. They found dorsal side kernel orientation and selection of specific wavelength regions had significant impact in detecting DHV wheat. Bleached kernels had lower classification accuracies (91.1 – 97.1%) than non bleached kernels (97.5 – 100%) in the two class PLSR model. Negative beta coefficients at the wavelength region of 550 – 650 nm were responsible for wheat kernel color. They found wavelengths in Vis/NIR region or NIR region alone were more suitable than in visible region for the detection of DHV wheat kernels from NDHV wheat kernels.

Wang et al. (2002) studied the use of NIR spectroscopy to differentiate sound soybean seeds from damaged soybean seeds. Six categories of soybean seeds (sound kernels, weather damaged kernels, frost damaged kernels, sprout damaged kernels, heat damaged kernels, and mold damaged kernels) were taken for this study. A diode array NIR spectrometer was used to collect the NIR reflectance intensities of samples in the wavelength region of 400 – 1700 nm at 5 nm intervals. Two class and six class PLSR models were developed using commercial PLS software. Two class and six class ANN models were developed using back propagation neural network (BPNN) architecture to differentiate healthy soybean kernels from damaged soybean kernels. Two class PLSR model at NIR wavelength (750 – 1690 nm) and full wavelength (490 – 1690 nm) regions gave more accurate results for the calibration set of samples (> 99.3%) as well as the validation set of samples (> 99.5%) in discriminating the damaged kernels from the sound kernels than that of the PLSR model at the visible wavelength region (490 – 750 nm) (> 98.4% for the calibration set, > 97.8% for the validation set). Six class PLSR model at full wavelength region (490 – 1690 nm) produced good results in classifying sound and damaged soybean seeds for both the calibration (average classification accuracy = 75.2%) and the validation (average classification accuracy = 74.5%) set of samples.

NIR reflectance spectroscopy was used to identify heat damaged wheat kernels from sound kernels (Wang et al. 2001). Heat damage of wheat was measured using the mixogram and the measured data were used as the reference data set. Gelatinization, pasting, and set back profiles of wheat were measured using a rapid viscosity analyzer. Diode array spectrometer was used for measuring the Vis/NIR reflectance intensities

from heat damaged and undamaged kernel samples in the wavelength region of 400 – 1700 nm. PLSR and two wavelength regression methods were used for model development to discriminate heat damaged wheat kernels from undamaged wheat kernels. Heat damaged wheat kernels were darker and more yellow in color than undamaged wheat kernels. A seven factor PLSR model produced a classification accuracy of 100% at the NIR wavelength region (750 – 1700 nm) for classifying heat damaged wheat kernels from undamaged wheat kernels. A two wavelength regression model (985 nm and 1050 nm) produced the best results for the calibration set (classification accuracy = 97.5%) and the test set (classification accuracy = 96.8%) of wheat samples. NIR spectroscopy could be effectively used in identifying heat damaged wheat kernels from healthy wheat kernels.

2.3.2. Chemical quality assessment of agricultural commodities using NIR spectroscopy

Protein content in wheat is affected by various environmental, agricultural, and genetic factors (Fig. 2.2).

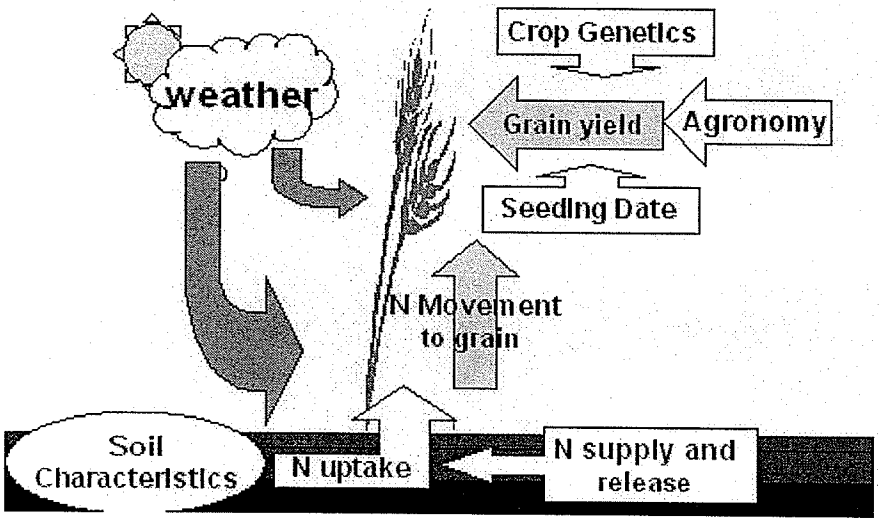

Fig.2. 2 Factors affecting protein content in wheat.
Source: Anonymous (2007c).

NIR reflectance spectroscopy was used for measuring the protein content in single kernels of wheat (Delwiche 1998). NIR wavelengths regions of 1100 – 2498 nm and 1100 – 1798 nm were used for scanning samples at 2 nm intervals and modeling purposes, respectively. Protein content of single kernels of wheat was determined using a combustion method and used as the reference data set. PLSR and MLR models were developed for protein prediction in single kernels of wheat. In this study, 3180 wheat kernels were taken (10 kernels each from 318 wheat samples) from two types of hard wheat and soft wheat varieties. The calibration, validation, and test sets of samples were developed for protein modeling. In addition to that, red and white wheat varieties were pooled together, and pooled RED and WHITE calibration, validation, and test sets were formed. Finally, all red and white wheat varieties were pooled and the three sets were

formed. In order to improve the PLSR model performance, MSC was used as a data preprocessing tool. In overall error analysis, errors related to reference method, scan repeatability, and chemometric studies were found and added together. The relationship between kernel size and wheat protein could not be developed. MSC effectively reduced the spectral variability. NIR wavelength region of 1100 – 1400 nm was found effective for protein content prediction analysis. In error analysis, chemometric error was equal to 0.411% protein. Although stepwise MLR models produced the test set standard error of prediction (SEP) values of 0.1 – 0.2% protein more than that of PLSR models, stepwise MLR was recommended for protein prediction because of its simplicity. It was recommended to include wheat samples from more than one crop year in calibration to improve the prediction accuracy of the models.

Wesley et al. (2001) developed an NIR model to measure the gliadin and glutenin contents of wheat. Gliadin and glutenin contents are directly related to the quality of wheat protein. PLSR with one-out cross validation and curve fitting methods were used to predict gliadin and glutenin of wheat from NIR spectral data. Size exclusion HPLC was used for measuring protein content in wheat flour samples and used as the reference data set. Preprocessing of NIR spectral data was done using SNV-DET for scatter correction. It was observed that the performance of PLSR method was better than that of curve fitting method in measuring glutenin and gliadin contents of wheat. PLSR method had coefficient of determination (R^2) values of 0.83 and 0.78 for glutenin and gliadin contents, respectively, whereas curve fitting method had R^2 values of 0.71 and 0.46 for glutenin and gliadin contents, respectively. Standard error or cross validation (SECV) values for glutenin and gliadin contents using PLSR method were 0.38 and 0.43,

respectively, whereas SEP values of curve fitting method for glutenin and gliadin contents were 0.65 and 1.02, respectively. The curve fitting method had the potential to rank the samples qualitatively as high, medium, and low protein samples based on the presence of glutenin and gliadin contents.

Ruan et al. (2002) developed an ANN model to measure the mycotoxin DON in barley using NIR spectroscopy. Barley samples from different crop years and different levels of mold damage were used for this study. DON concentration values of barley samples were measured using gas chromatography / mass spectrometry and used as the reference data set. Wavelengths of 400 – 2500 nm at 2 nm intervals was used to collect the absorbance values from bulk barley samples. Barley samples of different DON levels were clearly discriminated in NIR wavelength region of 1500 – 1800 nm. ANN models were developed using three layer BPNN architecture. Raw NIR absorbance values were used as input for ANN model development. Effect of wavelength increments and wavelength ranges were tested for the improvement in DON prediction in barley. Wavelengths in the visible region (400 – 700 nm) and short wavelength NIR region (700 – 1100 nm) found to be crucial in predicting the DON levels in barley. ANN models in the visible region (400 – 700 nm) and the short wavelength NIR region (700 – 1100 m) produced the best R^2 values of 0.921 and 0.912, respectively; and produced the minimum SEP values of 3.351 and 3.706, respectively. ANN models with NIR absorbance values from 400 – 2500 nm at 2 nm and 4 nm intervals produced the best R^2 values of 0.933 and 0.923, respectively; and produced the minimum SEP values of 3.097 and 3.431, respectively.

Feasibility study to measure the protein in wheat kernels using NIR transmittance spectroscopy was done by Delwiche (1995). Five samples each from six wheat classes of different protein levels were used for this study. Ninety eight wheat kernels were randomly chosen in each sample. Protein content in each test sample was found using nitrogen and food protein determinator; and Kjeldal method, used as the reference data set. NIR transmittance values were taken in the wavelength region of 740 – 1139 nm and converted into absorbance using $\log_{10}$ (1/T). NIR absorbance features in the wavelength region of 850 – 1050 nm were used for PLSR model development. Four sample sets were used for model calibration and one set for model validation. Three types of input data such as i) no change in log (1/T) value (absorbance value), ii) Multiplicative scatter correction (MSC) done in absorbance values, and iii) MSC done in second derivative of log (1/T) value, were used. It was observed that model performance was improved using the second and third data pretreatment methods.

Wang et al. (2004) developed a linear calibration model for determining moisture in ground wheat using NIR spectroscopy. NIR spectral data were collected in the wavelength region of 850 – 2000 nm at 5 nm intervals. NIR reflectance data were then converted into absolute NIR absorbance using A = log (1/R). Linear calibration models were developed for the averaged NIR spectra and the first derivatives of averaged NIR spectra. Models were also developed using MSC and SNV preprocessed NIR absorbance data set. It was observed that the first derivatives of averaged NIR spectra gave better results in linear calibration modeling (R^2 = 0.972, RMSEC = 0.239). Baseline elimination and resolution of overlapping peaks were found helpful in improving the performance of the model developed from the first derivatives of NIR absorbance values. R^2 and RMSEC

values of the model developed from the averaged NIR spectra were 0.793 and 0.541, respectively.

An NIR model was developed to measure the starch structure and the degree of processing in cereal products produced from twin screw extrusion cooking (Guy et al. 1996). Non invasive measurements of hot melt in the extruder barrel were predicted using NIR spectroscopy. NIR reflectance intensities were collected in the wavelength region of 1100 – 2500 nm at 4 nm intervals. NIR models were developed using forward stepwise regression method to predict the specific mechanical energy (SME) using the principal component scores. Principal components were identified for the NIR reflectance intensities in the wavelength region of 1300 – 1800 nm using PCA. The first two principal components could explain > 95% of variation of the NIR reflectance intensities. The damage of hydrogen bonds between hydroxyl groups of the starch molecules at the time of extrusion cooking affected the performance of the models.

Quality parameters of wheat flour were determined using NIR models (Miralbes 2004). AACC standard methods were used for measuring the quality parameters such as moisture content, protein, wet gluten, dry gluten, and ash content in wheat. NIR transmittance data were collected in the wavelength region of 850 – 1048.2 nm at 2 nm intervals. SNV-DET and modified PLSR were used for correcting the spectral variations of the NIR transmittance data and developing the calibration models, respectively. NIR absorbance data were treated with SNV-DET and first derivative processing methods, and used as input for PLSR model development. The PLSR model produced the best R^2 values of 0.99 for the validation sets to predict protein and moisture contents in wheat flour. It also produced the minimum SEP of 0.14 and 0.15 in predicting protein and

moisture contents of wheat flour, respectively. NIR transmittance spectroscopy could be effective in controlling the product quality online in milling industries.

Petterson and Aberg (2003) determined DON mycotoxin in wheat kernels using NIR transmittance spectroscopy. NIR transmittance spectroscopy was used to study the level of ergosterol and DON in wheat. Insect and mite infestations were also studied from NIR transmittance spectra of wheat. Wavelength region of 570 – 1100 nm was used for acquiring NIR transmittance intensities. Gas chromatography (GC) and high performance liquid chromatography (HPLC) were used to measure the DON levels of wheat and these data were used as the reference data set. PLSR models were developed for determining DON levels of wheat. Eleven to thirteen PLS factors were extracted based on the input data derived from three different wavelength regions (normal (850 – 1100 nm), extended (570 – 1100 nm), and reduced (670 – 1100 nm)) for model development. PLSR models were developed for fungal infected wheat samples grown in Norway and Austria separately. Eleven factor PLSR model for Nordic sample produced the best r value of 0.984 and SECV of 381μg DON per kg of wheat at a reduced wavelength region of 670 – 1100 nm. NIR transmittance spectroscopy could be a potential technique to measure DON and another mycotoxin, ergosterol in cereal crops.

2.3.3 Quality assessment of agricultural commodities using NIR spectroscopy

Delwiche (2003) identified scab and mold damage in wheat kernels using NIR reflectance spectroscopy. Mold-affected and scab-damaged kernels in HRS wheat were separated from sound kernels using visual inspection. A Zeiss MCS511 diode array spectrometer was used to collect NIR absorbance values in the wavelength region of 940 – 1700 nm. Models were developed using the NIR absorbance values of 1002 – 1704 nm

at 6 nm interval as input. Two types of kernel orientation techniques (crease down placement and random placement of kernels) were used. LDA with leave-one-out cross validation, soft independent modeling of class analogy (SIMCA), and PCA were used for statistical purposes. Cross validation accuracies of 89 – 98% and test set accuracies of 90.5 – 98.4% were observed in the two way classification of sound kernels and damaged kernels (scab damaged + mold damaged) using LDA. LDA with 1, 2, 7, and 3 principal component scores produced the cross validation and the test set accuracies of 89.3% and 86.4%, respectively. Cross validation and test set accuracies ranged from 85.3 – 86.7% and 83.6 – 85.8%, respectively, in 2, 4, and 6 factor SIMCA-PLS models. Classification accuracies of a test set of scab damaged, mold damaged, and sound kernels with random kernel orientation were equal to or less than that of precise kernel orientation.

Baker et al. (1999) developed a NIR model to differentiate the kernels infested by larval and pupal stages of rice weevil (*Sitophilus oryzae* (L.)) which was parasitized by a mite, *Anisopteromalus calandrae* (Howard), from uninfested kernels and unparasitized kernels of wheat. Uninfested kernels, kernels infested with weevil larvae, kernels infested with weevil pupae, kernels containing parasitoid larvae, and kernels containing parasitoid pupae were used for this study. A diode array NIR spectrometer was used to collect NIR absorbance values of wheat kernels in the wavelength region of 400 – 1700 nm. A PLSR model was developed using thirteen PLS factors to detect the rice weevil infestation in wheat kernels. Thirteen factor PLSR model produced the best r value of 0.90 and the minimum SECV value of 0.15. NIR spectroscopy was effective in this study because of the different levels of NIR absorption due to the compositional difference in chitin and cuticle contents of insects. Misclassifications occurred in the kernels with small

parasitoid or weevil that absorbed a small amount of NIR radiation. NIR spectroscopy was effective in differentiating the larval and pupal stages of insect or parasitoid in wheat kernels as they absorbed a threshold amount of NIR radiation.

Feasibility of NIR spectroscopy in detecting the insect fragments in wheat flour was studied by Perez-Mendoza et al. (2003). A diode array NIR spectrometer was used for collecting the diffuse reflectance intensities in the wavelength region of 550 – 1700 nm from wheat flour samples. These intensities were then converted into NIR absorbance values. AOAC 972.32 floatation method was used as the reference method to find the insect fragments in wheat flour samples. A PLSR model was developed using ten PLS extracted factors to predict the insect fragment levels in wheat flour. Six wavelengths (890, 1120, 1220, 1370, 1530, and 1630 nm) in the NIR region were identified as the wavelengths related to the excitation of the first, second, and third overtones of CH groups. Some of the main constituents of insect fragments such as chitin and lipid were responsible for CH group absorption of NIR radiation. The AOAC floatation method had high accuracy in predicting the insect fragments in wheat flour samples and took 2 h for determination. Classification accuracy of 83.3% was reported in classifying samples with 0, 35, and 75 insect fragments as < 130 fragment sample per 50 g flour class; and of 90% for the samples with 150, and 300 insect fragments as > 130 fragment sample per 50 g flour class. The NIR spectroscopy was not as sensitive as the floatation method in determining the US-Food and Drug Administration (US-FDA) level of insect fragments (75 insect fragments per 50 g of flour) in wheat flour. But there was a possibility for NIR and mid-IR spectroscopy to detect the US-FDA permissible levels of insect fragments in wheat flour with future advancements in the field of spectroscopy.

Wang et al. (2004) used NIR spectroscopy to classify fungal damaged kernels from healthy soybean kernels. PLSR and ANN methods were used for classification purposes. Healthy soybean kernels and soybean kernels damaged by *Phomopsis*, *Cercospora kikuchii*, soybean mosaic virus (SMV), and downy mildew were used. A diode array spectrometer was used for collecting the NIR reflectance intensities in the wavelength region of 400 – 1700 nm from single kernels of soybean. NIR reflectance intensities were interpolated to 5 nm intervals. Five hundred healthy seeds and eight hundred fungal damaged seeds were used for this study. A two class model (healthy vs. damaged) and five class model (healthy vs. four types of damage) were developed using PLSR and ANN methods and NIR reflectance intensities in the wavelength region of 490 – 1690 nm. A ten factor PLSR model produced classification accuracies of > 99% for the calibration and validation sets in classifying healthy and damaged soybean kernels. A five class ANN model produced the highest average classification accuracy of 93.5% for the calibration set and 94.6% for the validation set in the wavelength region of 490 – 1690 nm in classifying healthy and four fungal damaged soybean kernels separately.

Detection of the mycotoxin fumonisin in corn using the reflectance and transmittance spectroscopy at Vis/NIR wavelengths was studied by Dowell et al. (2002). Two different spectrometers, one for the transmittance mode (Fiber optic spectrometer, Model S2000, Ocean Optics, Dunedin, FL) and the other for the reflectance mode (NIR spectrometer, Perten Instruments, Springfield, IL), were used for collecting the transmittance intensities from 550 – 1050 nm and reflectance intensities from 400 – 1700 nm from single kernels of corn. Total fumonisin in single kernel of corn was found using a fluorometer and was used as the reference data set. A PLSR method was used to

develop a calibration model to detect the fumonisin content (< 10 ppm and ≥ 10 ppm) in corn. Mahalanobis distance method was used for grouping the kernels based on fumonisin levels. Classification errors were 0% and < 7.2% in identifying 1 – 10 ppm and > 100 ppm of fumonisin, respectively in corn. Misclassification of kernels was more for corn with 10 – 100 ppm of fumonisin (error rate = 23.5 – 73.0%). Wavelengths of 650, 710, 935, and 990 nm were identified as the important transmittance wavelengths using PLS beta coefficients. Also, 590, 995, 1200, and 1410 nm wavelengths were identified by the PLS beta coefficients that these wavelengths were important reflectance wavelengths for classification. Fumonisin in corn kernels did not absorb a considerable amount of NIR radiation. Changes in chemical compositions of corn due to fumonisin levels might influence the NIR absorption levels. The possibility of the detection of fumonisin at a minimum FDA allowed threshold level (2 – 4 ppm) in corn using NIR transmittance or reflectance spectroscopy is questionable.

Pearson et al. (2001) studied the detection of the mycotoxin aflatoxin in corn using Vis/NIR transmittance and reflectance spectroscopy. Corn samples were obtained at four different aflatoxin levels using black light examination on bright greenish-yellow fluorescence (BGYF) characteristics. A silicon photo diode array fiber optic spectrometer was used for collecting the transmittance spectra of single corn kernel samples in the wavelength region of 500 – 950 nm. Diode array NIR spectrometer was used for collecting the reflectance intensities of corn in the wavelength region of 550 – 1700 nm. Nineteen point Savitzky-Golay second order filtering operation was used for smoothing the transmittance spectra. Mahalanobis distance method was used for grouping the corn samples of three different aflatoxin levels: 1, 10, and 100 ppb. A PLSR method was used

for model development to detect aflatoxin in corn. Aflatoxin in single kernels of corn was found using USDA-FGIS aflatest affinity chromatography and used as the reference data set. Transmittance and reflectance data produced good classification accuracies with the error rates of < 5% in identifying 0, 1 – 10, > 100 ppb aflatoxin in corn using discriminant analysis. Error rates were of 52.2 – 56.5% and 65.2 – 87.0% for transmittance and reflectance data, respectively, in classifying 10 – 100 ppb aflatoxin in corn using discriminant analysis. Germ up and germ down kernel orientation improved the classification accuracies of 10 – 100 ppb aflatoxin in corn to ≥ 84% using six factor PLSR model developed from NIR reflectance intensities. Vis/NIR spectroscopy could be used to detect aflatoxin in corn.

Maghirang et al. (2003) studied the detection of live or dead rice weevil at its pupal and larval stage in single kernels of wheat using NIR spectroscopy. Sound kernels, kernels with pupae, large larvae, medium sized larvae, and small larvae of rice weevil were selected using x-ray imaging. Spectral data of all the above kernel samples with live and dead internal insects were collected and analyzed. A single kernel characterization system (SKCS) was used to collect the spectral data of individual kernels in the wavelength region of 400 – 1700 nm at 5 nm intervals. A PLSR method was used to develop a model at the wavelength region of 950 – 1690 nm for internal insect detection in wheat. PLS beta coefficients identified ten important wavelengths (990, 1135, 1210, 1250, 1370, 1395, 1425, 1510, 1610, and 1670 nm) responsible for detecting internal insects in wheat. Classification accuracies were > 90% for detecting wheat samples with pupae or large larvae of rice weevil using 5 – 7 factor PLSR models.

2.4 Quality assessment of agricultural commodities using NIR hyperspectral imaging

Cogdill et al. (2004) developed NIR models for predicting moisture and oil contents in maize using PLSR, PCR and genetic algorithm with MLR. They used optical absorbance data as input for these models. Optical absorbance values were obtained from standardized NIR transmittance intensities collected in the wavelength region of 750 – 1090 nm at 5 nm intervals. These NIR transmittance data were standardized using an opal glass standard. Data pretreatments were done using SNV, DET, and MSC. An eleven factor PLSR model produced the minimum SECV of 1.20% m.c. and relative performance determinant (RPD) of 2.74 for moisture prediction in maize using optical NIR absorbance data after the removal of outliers. A nine factor PLSR model produced the minimum SECV value of 1.38% oil and RPD value of 1.45 for oil content prediction in maize using SNV pretreated, outlier removed, optical NIR absorbance data.

Wang and Paliwal (2005) studied the performance of a morphological shrinking algorithm and simulated ellipsoidal surface fitting in reducing the spectral variability of different parts of the wheat kernel due to its curved surface. The difference in the reflectance levels of the curved surfaces of the agricultural products was the main cause for spectral variability. Hyperspectral images of four western Canadian wheat classes were acquired in the wavelength region of 1100 – 1600 nm at 30 evenly spaced slices. It was found that a morphological shrinking method was more effective in correcting the overall spectral variability than that of ellipsoidal surface fitting for all classes of wheat. Both the methods performed well in reducing the spectral variability along the minor axis

of the endosperm of wheat kernels. Ellipsoidal surface fitting was not effective in correcting the spectral variability along the major axis of the endosperm of wheat kernels.

Nagata et al. (2005) developed stepwise MLR models for measuring firmness and soluble solids content (SSC) in strawberries. NIR hyperspectral images of strawberries were collected in the wavelength region of 650 – 1000 nm at 5 nm intervals. A three wavelength (685, 985, and 865 nm) MLR model produced correlation of 78.6% and SEP of 0.350 in predicting firmness in 50% to a full-ripe group of strawberries. Correlation of 87% and SEP of 0.43 were reported from five wavelength (915, 765, 870, 695, and 860 nm) MLR model in predicting SSC in 70% to the full-ripe group of strawberries. These results confirmed that 675 and 980 nm wavelengths were responsible for chlorophyll and water absorption, respectively. Carbohydrate and sugar absorptions were at the NIR wavelength of > 800 nm.

Bruises were detected in apples using NIR hyperspectral imaging in the wavelength region of 900 – 1700 nm (Lu 2003). NIR region of 1000 – 1340 nm produced the best results in bruise detection in apples. Two classes of apples (Red Delicious and Golden Delicious) were used for this study. Bruises were created in apples by mechanical impacts. A pendulum fixed at 127, 229, and 330 mm from the equatorial line of an apple was used to create bruises by impact. Feature extraction from NIR hyperspectral images was done using principal components (PC) and minimum noise fraction (MNF) transforms. Bruise detection accuracies were 62 – 88% for Red Delicious apples and 59 – 91% for Golden Delicious apples to detect old and new bruises using both the transforms. The optimal spectral resolution and corresponding number of spectral bands were 8.6 – 17.3 nm and 20 – 40, respectively.

Mehl et al. (2004) detected surface defects and contaminations in apples using hyperspectral imaging with a high spatial resolution of 0.5 – 1 nm. Several defects and contaminations (side rots, bruises, flyspecks, scabs, molds, fungal diseases, and soil contaminations in apples) were detected in this study. Four classes of apples (Red Delicious, Golden Delicious, Gala, and Fuji) were used. Monochromatic images of apples were taken in the wavelength region of 430 – 900 nm. Asymmetric and symmetric second difference models were developed to analyze multispectral images for sorting out good apples from the contaminated ones. Chlorophyll absorption wavelength of 685 nm and two NIR wavelengths of 722 nm and 865 nm were found responsible for detecting defects and contaminations in apples irrespective of their color and cultivar.

Bitter pit lesion is created in apples due to a physiological disorder. This disorder creates brown colored lesions under the epidermis of apples. It reduces the consumption of apples and sometimes leads to rejection of the apple lot as a whole by exporters. NIR hyperspectral imaging was used for identifying bitter pit lesions on apples (Nicolai et al. 2006). A discriminant PLS model with leave-one-out cross validation was developed to find bitter pit lesions on apples and the model was further validated. PLS calibration was done using two latent variables. Bitter pit lesions were identified in PLS predicted images and binary images of apples. NIR hyperspectral images could detect bitter pit lesions on apples. But, this method could not differentiate bitter pit lesions and corky tissues developed inside apples.

Summary

There is a scope for using NIR hyperspectral imaging to classify western Canadian wheat classes. Moisture levels can affect identification and differentiation of

the wheat classes using NIR absorbance values as input. Five different levels of moisture (straight (12% and 14%), tough (16%), and damp (18% and 20%)) could be used for the identification of wheat classes. In NIR spectroscopy, statistical and neural network classifiers produced good results in identifying different types of damages (Wang et al. 2002) and different types of fungal damages (Wang et al. 2004) in soybean, identifying waxy wheat (Delwiche and Graybosch 2002), determining DON levels in barley (Ruan et al. 2002), and detecting fumonisin levels (Dowell et al. 2002) and aflatoxin levels (Pearson et al. 2001) in corn. Wavelength region of 960 – 1700 nm and its subset of 1100 – 1600 nm can be used for wheat class identification.

It is also observed that there is scope for predicting the protein and oil contents of wheat using NIR hyperspectral imaging with NIR absorbance values as input. There were no studies done in the prediction of protein and oil contents of wheat using the reflectance or absorbance values of NIR hyperspectral imaging. This study will prove the ability of NIR hyperspectral imaging in predicting the major constituents of wheat. Wavelength region of 960 – 1700 nm and its subset of 1100 – 1600 nm can be used for the development of NIR prediction models. In NIR spectroscopy, PLSR was used for the model development to determine protein (Delwiche 1998), to assess heat damage (Wang et al. 2001), to detect live or dead internal rice weevil (Maghirang et al. 2003) in wheat. A PLSR method is chosen in this study for developing protein and oil content prediction models in wheat. Summary of different methods with important features used, accuracies attained, statistical tools used, and wavelength regions and intervals used are shown in Appendix (Tables A.1 and A.2).

3. MATERIALS AND METHODS

Different materials and methods used in this NIR hyperspectral imaging study on wheat for class identification and protein and oil content prediction are explained in this section.

This study is divided into three segments:

1. Differentiation of eight western Canadian wheat classes at same moisture level (Study: 1)
2. Differentiation of five western Canadian wheat classes at five moisture levels with and without moisture effects (Study: 2)
3. Prediction of protein and oil content in wheat for five western Canadian wheat classes at five moisture levels (Study: 3)

The major components involved in completing these studies are NIR hyperspectral imaging system, preparation of grain samples, laboratory measurement of protein and oil content in wheat, NIR hyperspectral images handling, and analysis of NIR hyperspectral data.

3.1 NIR hyperspectral imaging system (Studies: 1, 2, and 3)

The NIR hyperspectral imaging system used in this study was a long wavelength NIR camera with a liquid crystal tunable filter, lens, sample stage, and light source controlled through a Dell Optiplex GX280 Intel(R) (Dell Inc., Round Rock, TX) computer (Fig. 1). An Indium Gallium Arsenide (InGaAs) camera (Model No. SU640-1.7RT-D, Sensors Unlimited Inc., Princeton, NJ) was used for acquiring images at different wavelengths in the NIR region of 960 – 1700 nm. The camera could be operated

in a room at a temperature range of 20 to 40°C and had 12 bit digitization capacity. This system has spatial resolution of 640 × 480 pixels with 27 μm pitch.

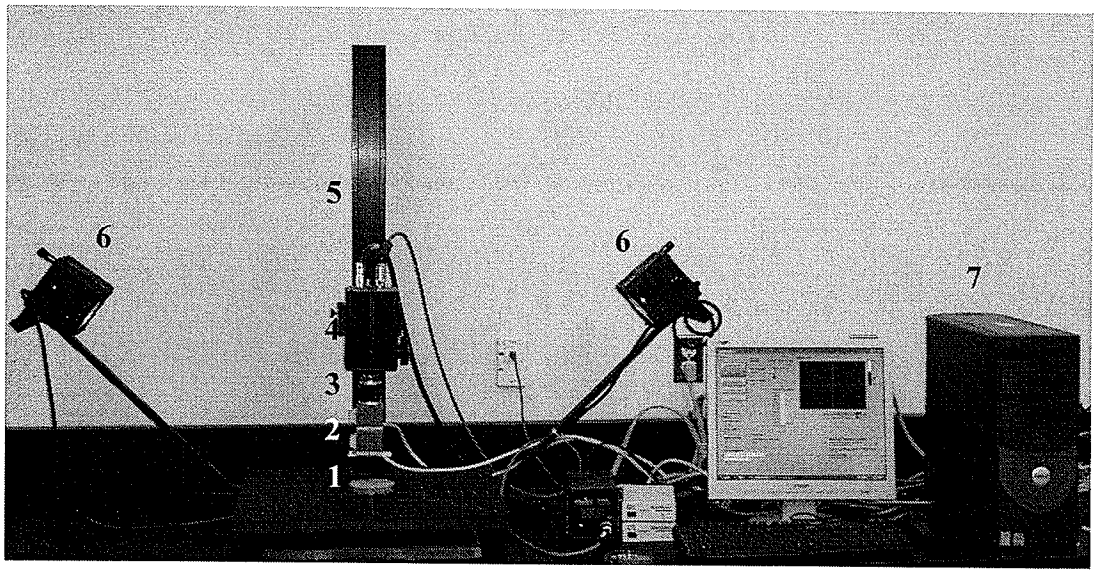

Fig. 3.1 NIR hyperspectral imaging system.

1. Bulk wheat sample, 2. Liquid crystal tunable filter (LCTF), 3. Lens, 4. NIR camera, 5. Copy stand, 6. Illumination, and 7. Data processing system.

The electronically tunable, liquid crystal tunable filter (LCTF) (VariSpec model No. MIR 06, Cambridge Research and Instrumentation Inc., Woburn, MA) had 20 mm aperture and 10 mm transmission bandwidth. This high quality interference filter helped to rapidly select a wavelength in the NIR region without any vibration. This filter system was attached to the camera system which ultimately helped to acquire multispectral images with 16 bits ultra high resolution. The data acquisition board (NI PCI-1422, National Instruments Corporation, Austin, TX) was attuned to RS-422 signals generated from the camera system for image acquisition.

The sample was illuminated by a pair of 300 W halogen bulbs (USHIO Inc., Chiyoda-ku, Tokyo, Japan) fitted on either side of the copy stand of the NIR imaging

system. These halogen bulbs had the capacity to emit light in a wavelength range of 400 – 2500 nm. Halogen regenerative cycle was developed inside the bulb that helped to prevent the blackening of bulb wall and avoid the reduction in light output (Anonymous 2007d). NIR hyperspectral images were acquired with the help of a control program written in LabVIEW (Version 1, National Instruments Corporation, Austin, TX). Dark current image was taken every time by blocking the entrance slit using a black board to remove the inherent noise in the system before acquiring NIR hyperspectral images of wheat samples. The camera was aligned to the centre wavelength of 1330 nm in the NIR camera's usable wavelength region of 960 – 1700 nm. Dark current subtracted NIR hyperspectral images were automatically taken and saved using the control program of the camera. Spectral normalization was done to remove the effect of variations in illumination at the time of image acquisition.

3.2 Preparation of grain samples

3.2.1. Grain samples for wheat class identification (Study: 1)

Wheat samples from eight classes, i.e., CWRS, CPSR, CWES, CWRW, CPSW, CWAD, CWSWS, and CWHWS were obtained from seed distributors in Manitoba and Alberta, two major wheat producing provinces in Canada. Samples for each wheat class were combined into a single mixed sample for each class. In Canada, wheat of one specific class from different growing regions is not segregated and is marketed as a composite class. The Canadian grain handling and transportation system causes the grain of the same grade grown in various parts of Canada to mix thoroughly by the time it reaches the terminal elevator (CWB 2007). Wheat for each class was conditioned to a

moisture level of 11% (wet basis). Moisture contents of wheat were determined using ASAE S352.2 standard method (ASAE 2003).

3.2.2 Grain samples for class differentiation with and without moisture effects; and protein and oil content prediction in wheat (Studies: 2 and 3)

Five wheat classes, i.e., CWRS, CWES, CWRW, CWSWS, and CWHWS were obtained from seed distributors in Manitoba and Saskatchewan. About a 10 kg composite sample in each wheat class was conditioned to five different moisture levels (12%, 14%, 16%, 18%, and 20% wet basis).

3.3 Laboratory measurement of protein and oil content in wheat (Study: 3)

3.3.1 Protein

Wheat samples for protein content were analyzed by the staff at the Norwest Labs, Winnipeg, MB. In their lab, protein in wheat was determined by measuring total nitrogen by combustion using AOAC official method 990.03 for four replicates per wheat class (AOAC 2003).

3.3.2 Oil content

Oil content in wheat was measured at the Canadian Wheat Board Centre for Grain Storage Research, University of Manitoba, Winnipeg. It was determined using AACC standard method 30-25 for three replicates per wheat class (AACC 2000). Wheat samples were dried in hot air oven at 130°C for 19 h and ground using a mill (M-2, Seedburo Equipment Co., Chicago, IL). Ground wheat samples (5 g each) were taken in a filter paper and placed inside the extraction thimble. Empty weights of the extraction beakers were measured and petroleum ether (30 ml each) was taken. Extraction thimbles and

extraction beakers were fixed to the extractor unit and ether was allowed to circulate in the system using condenser unit. The entire experiment was allowed to run for 6 h. Ether passed through the extraction thimbles and extracted the oil from the sample. The oil and solvent mixture obtained at the end were separated by evaporating the solvent. Amount of oil collected was worked out by subtracting the empty weight of the beaker from the weight of the beaker with oil. Crude oil content was measured using the following formula

$$\text{Crude fat or ether extract, \%} = \frac{\text{Weight of oil content} - \text{blank}}{\text{Weight of sample}} \times 100$$

3.4 NIR hyperspectral image handling (Studies: 1, 2, and 3)

NIR hyperspectral images of bulk wheat samples were collected at the equipment's usable wavelength range of 960 – 1700 nm incremented by 10 nm. The NIR wavelength region was segmented into 75 slices, resulting in the NIR hyperspectral image cube having 75 images (an image per slice) in it with the first image at 960 nm. The camera was focused at a wavelength of 1330 nm. NIR reflectance intensities were recorded at each wavelength slice of a hypercube depending on the NIR absorbance properties of the main constituents (protein, oil, starch, and moisture content) of wheat. NIR reflectance intensity of a sample will be high when the sample absorbs a low amount of NIR radiation and vice versa.

3.4.1 NIR hyperspectral image acquisition (Study: 1)

Wheat samples were prepared by randomly taking sub samples of 50 g each from the composite class sample and placing in a petri dish of 90 mm diameter and 11 mm depth. Three hundred NIR hyperspectral images were collected for each wheat class. In total, 75 × 300 × 8 (180,000) images were taken and analyzed.

3.4.2 NIR hyperspectral image cropping and feature extraction (Study: 1)

Cropping at the center of the image was done to avoid pixels with poor reflectance values along the four edges of the image. A pixel at the spatial coordinates of 190, 325 was identified as the center of the NIR hyperspectral image. An area of 200 × 200 pixels around the center pixel, i.e., 100 pixels in all four directions from the center pixel (top, bottom, left, and right), was cropped from each image and the reflectance value of each pixel of the cropped region at each slice of the NIR hyperspectral image was extracted. Single median spectrum was calculated from a region of interest of a hyperspectral image and used for developing PLS calibration models (Burger and Geladi 2006). Liu et al. (2007) used mean spectra of sub images derived from the hyperspectral images and used them for determining SSC of oranges using PCA and line regression methods.

Mean reflectance and normalized mean reflectance values were calculated for all wavelength slices of the NIR hyperspectral images. The mean reflectance values were normalized at each wavelength slice of an NIR hyperspectral image by dividing reflectance at each slice by the maximum reflectance of all slices in an NIR hyperspectral image. Normalized mean reflectance values at each wavelength slice of an NIR hyperspectral image was considered as a feature for developing a model to differentiate

wheat classes. In total, 75 features were extracted from an NIR hyperspectral image of a bulk wheat sample. Image cropping and feature extraction were done in MATLAB (Version 7, The Mathworks, Inc., Natick, MA).

3.4.3 Class differentiation with and without moisture effects; and protein and oil content prediction in wheat (Studies: 2 and 3)

3.4.3.1 NIR hyperspectral image acquisition

Wheat samples were prepared by randomly taking a petri dish full of sub samples from the composite class sample. The top surface of wheat bulk sample was made flat by scraping in two perpendicular directions with a ruler. One hundred NIR hyperspectral images were collected for each wheat class at each moisture level. In total, 75 × 100 × 5 × 5 (187,500) images were taken and analyzed.

3.4.3.2 NIR hyperspectral image cropping and feature extraction

Cropping at the center of the image was done to avoid pixels with poor reflectance values along the four edges of the image. A pixel at the spatial coordinates of 190, 325 was identified as the center of the NIR hyperspectral image. An area of 200 × 200 pixels around the center pixel, i.e., 100 pixels in all four directions from the center pixel (top, bottom, left, and right), was cropped from each image and the reflectance value of each pixel of the cropped region at each slice of the NIR hyperspectral image was extracted.

Mean reflectance values were calculated for all wavelength slices of the NIR hyperspectral images. Normalization of reflectance was done to avoid the variations in

light intensity while acquiring NIR hyperspectral images. Absorbance at each wavelength slice of an NIR hyperspectral image was calculated as:

$$A = \log_{10}(1/R)$$

where, A is the absorbance at a wavelength slice of an NIR hyperspectral image; R is normalized NIR reflectance at the same wavelength slice of NIR hyperspectral image. In total, 75 absorbance features were extracted from an NIR hyperspectral image of each bulk wheat sample.

3.5 NIR hyperspectral data analysis

3.5.1 Wheat class identification (Study: 1)

3.5.1.1 NIR reflectance spectra and their slopes

For each wheat class, an average of normalized NIR mean reflectance intensities at each wavelength slice of all NIR hyperspectral images was determined and used to plot NIR reflectance spectra. Slopes of the normalized NIR mean reflectance spectra were calculated as the rate of change in the normalized NIR mean reflectance intensity to the wavelength and were plotted against the wavelength. Slopes of normalized NIR mean reflectance intensities at different wavelengths are not the same and make the wavelengths to distinguish clearly from each other. Mohan et al. (2005) used percent NIR reflectance slopes and ratio of slopes as input features for classification of seven bulk cereal grains using Vis/NIR reflectance intensities and reported 99.5% of classification accuracy using linear parametric and neural network classifiers with top five slope features identified by STEPDISC procedure in SAS and BPNN architecture in ANN. In this study, change in trend in the average of normalized NIR mean reflectance spectra leads to the creation of peaks and troughs in the curve.

3.5.1.2 Classification

PROC DISCRIM (SAS 2002) was used for developing models using both LDA and QDA with leave-one-out cross validation. Normalized NIR mean reflectance intensities extracted from the NIR hyperspectral images of wheat samples were used as input features to these models. PROC STEPDISC was used to identify the top 10 wavelengths that contributed mainly to the classification of wheat classes. The level of contribution of each wavelength was identified by partial R^2 and average squared canonical correlation (ASCC) values. In STEPDISC, wavelengths with the highest level of contribution was identified and subsequently removed from further analysis to find the next best wavelength. This analysis was continued until the 10^{th} rank wavelength was found.

The ANN model was developed using a software package NeuroShell 2 (Ward Systems Group, Inc., Frederick, MD) to analyze the hyperspectral data. A modified-BPNN architecture with hidden layer having three slabs with different activation functions was used. Each slab of the hidden layer had a different activation function viz., Gaussian ($f(x) = e^{-x2}$), tanh ($f(x) = \tanh(x)$), and Gaussian complement ($f(x) = 1 - e^{-x2}$). As these slabs had different activation functions, different features of the input data set were created by the hidden layer to improve the performance of the ANN model. Output and input layers had 1 and 75 neurons, respectively. Each slab of the hidden layer had 25 neurons. The output slab had logistic activation function ($f(x) = 1 / (1+e^{-x})$). Input, hidden, and output layers of ANN had learning rate, momentum, and initial weight of 0.1, 0.1, and 0.3, respectively. ANN models for two different patterns i.e., 60% training, 30% test, and 10% validation sets (referred to as 60-30-10 model) and 70% training, 20% test,

and 10% validation sets (referred to as 70-20-10 models, respectively) were developed and results of these models were compared. Calibration interval (events) for a test set was fixed at 200. When the average error was < 0.0002 and epochs were > 1000 since the occurrence of the calculated minimum average error for training set, the training was stopped. Also, when the number of events was > 20000 since the occurrence of the calculated minimum average error for the test set, the training was stopped.

3.5.2 Differentiation of wheat classes with and without moisture effects (Study: 2)

PROC DISCRIM (SAS 2002) was used for developing models using LDA and QDA with leave-one-out cross validation. NIR absorbance features extracted from the NIR hyperspectral images of wheat samples in the wavelength regions of 960 – 1700 nm (75 slices) and 1100 – 1600 nm (51 slices) were used as inputs to these models. PROC STEPDISC was used to identify the top 10 wavelengths that contributed mainly to the classification of wheat classes at various moisture levels. The level of contribution of each wavelength was identified by the values of partial R^2 and ASCC. Classification accuracies of wheat classes were finally compared with the classification accuracies derived from LDA and QDA using the top 7 NIR absorbance features as input.

A modified-BPNN architecture with three slabs in the hidden layer was used for developing a 60-30-10 ANN model using the input features from 960 – 1700 nm (75 feature model), and from 1100 – 1600 nm (51 feature model). Calibration interval (events) for a test set was fixed at 200. When the average error was < 0.0002 and epochs were > 1000 since the occurrence of calculated minimum average error for the training set, the training was stopped. Also, when the number of events was > 20000 since it was the calculated minimum average error for the test set, the training was stopped.

Coefficient of determination (R^2), mean square error (MSE), mean absolute error (MAE), and correlation coefficient (r) were found for the training, test, and validation data sets. The top ten wavelengths were listed based on their contribution to ANN model for classification. Critical wavelengths of NIR absorbance spectra of wheat classes were identified using statistical and neural network classifiers with and without moisture effects. The classification accuracies of the training, test, and validation sets of the ANN model for wheat classes with and without moisture effects were reported.

PROC CANDISC was used for finding canonical variables using NIR absorbance values of wheat classes at different moisture levels as input. Top two canonical variables were plotted one against the other. The top canonical variable could explain the variation of a major component of the input sample. The second top canonical variable that had no correlation with the first canonical variable could explain the variation of the next major component of the input sample. This analysis was done to confirm the results of differentiation of wheat classes at different moisture levels.

Principal component analysis (PCA) was used to identify the wavelengths responsible for wheat class identification. Principal component (PC) scores, factor loadings, eigenvalues, Hotelling's T^2 distances were found for the mean reflectance intensities of NIR hyperspectral images of wheat classes at various moisture levels using MATLAB. The top ten principal components were identified using PROC STEPDISC in SAS. The top ten wavelengths were found based on their factor loadings for the top three principal components (2, 3, and 5).

3.5.3 Development of protein and oil content prediction model for wheat (Study: 3)

PROC PLS (SAS 2002) was used for developing PLSR models to predict protein and oil contents in wheat. For model validation, a test set validation method was used. NIR absorbance features extracted from the normalized NIR mean reflectance values of NIR hyperspectral images of wheat were used as input for PLSR models. Features from 60% of hyperspectral images in each wheat class at each moisture level were used as the training set, 30% as the test set, and 10% as the prediction set. Test set validation helped to reduce the number of extracted PLS factors to a minimum level at which the extracted factors had minimum root mean predicted error sum of squares (PRESS). The smallest number of factors where probability value exceeded 0.1 was selected for model development. PLSR model helped to explain the percent variation of independent and dependant variables using the extracted PLS factors. PLSR models for the prediction of protein and oil contents were developed for 1) NIR absorbance features at the wavelength region of 960 – 1700 nm (75 feature model) and 2) NIR absorbance features at the wavelength region of 1000 – 1600 nm (51 feature model).

Ratio of performance to deviation (RPD) is a statistic which is defined as the ratio of the standard deviation of measured protein or oil content to the standard error of predicted protein or oil content of wheat. It was calculated for PLSR models predicting protein or oil content in wheat. Measured protein or oil content with its standard deviation and predicted protein or oil content with its standard error were plotted. Scatter plot was prepared to show the variations in predicted protein or oil content and measured protein or oil content for all wheat classes. Predicted values of protein and oil content of wheat classes were grouped using Scheffe's test with $\alpha = 0.05$. Scheffe's test produces

high critical value in mean comparison and is considered as the most conservative method (Crow 2006).

3.5.4 Overview of different classification and prediction models used in studies 1, 2, and 3

3.5.4.1 Linear discriminant analysis (LDA)

Independent and dependent variables are called as features and classes of the model, respectively. They are expressed in measurement scale (independent variable) and nominal scale (dependent variable). Linear relationship among the features is developed to separate different classes. A line, plane, or hyperplane is formed based on the number of features involved in differentiating different classes. Covariances of features are assumed equal and the separating surfaces among classes are linear (Teknomo 2007).

3.5.4.2 Quadratic discriminant analysis (QDA)

Quadratic relationship among the features helps to separate different classes. Covariances of features are not equal and the separating surfaces among classes are quadratic (circle, parabola, or hyperbola) (Anonymous 2007e).

3.5.4.3 Artificial neural network (ANN)

An artificial neural network is a non-linear modeling tool, which can be used to develop a model from the input features through a network of nodes in input, hidden, and output layers to produce the output. Some of the important types of networks used in ANN are BPNN, Kohonen self organizing map network (Kohonen), probabilistic neural network (PNN), general regression neural network (GRNN), and group method of data

handling network (GMDH). BPNN, a supervised type of network in which both input and output variables contribute in training. Ward network, a modified form of BPNN, in which the hidden layer has multiple slabs and each slab has different activation functions was used in this study. This network is considered powerful as it forms different features from input nodes. Precision of ANN model is improved by selecting a proper network of input variables to explain the nature of the output variables (Neuroshell 2, Ward Systems Group, Inc., Frederick, MD).

3.5.4.4 Canonical discriminant analysis

Canonical discriminant analysis helps to reduce the dimensions of input variables like principal components analysis. Squared Mahalanobis distances between class means are found and both univariate and multivariate analyses of variance are performed in this method. Canonical variables plot helps to interpret the differences in output variables (groups) visually. In this method, canonical variables, otherwise orthogonal variables, are developed using linear combinations of input variables. The top most canonical variables have the highest possible multiple correlation with the output variables. The number of canonical variables developed is equal to the number of input variables used (Anonymous 1999a).

3.5.4.5 Partial least square regression (PLSR)

Partial least square regression model is developed using PLS factors found from linear combinations of the input variables. PLS factors are otherwise called as components or latent vectors. PLSR model has the capacity to explain both input and output variations. The main goal of PLSR is to develop a model by using a minimum

number of PLS factors to produce a minimum error in predicting the output (Anonymous 1999b).

3.5.4.6 Principal components analysis (PCA)

It is used to reduce the number of input variables to a minimum level for analysis using orthogonal linear transformation. In this method, input data set is transformed to a new coordinate system in such a way that the first principal component (first coordinate) explains the data set more effectively and attains the greatest variance. Similarly, second principal component of the model attains the second greatest variance in explaining the input data set. Generally, the top few principal components explain > 99% of variations of the input data set. The contribution of input variables in each principal component is found from the factor loading vectors (Anonymous 2007f).

4. RESULTS AND DISCUSSION

This section explains the results of different components of this study on wheat for class identification and protein and oil content prediction using the NIR reflectance and absorbance values of NIR hyperspectral imaging.

4.1 Wheat class identification (Study #: 1)

4.1.1 NIR reflectance spectra and their slopes

Normalized NIR mean reflectance spectra and their slopes of wheat classes are shown in Figs. 2 and 3, respectively. Normalized NIR mean reflectance intensities at 1400 – 1650 nm wavelength region showed separated spectral lines among various wheat classes. The slopes of NIR reflectance spectra resulted in corresponding peaks or troughs at 1580, 1490, 1280, 1100, 1610, and 1160 nm wavelengths. NIR absorption in different samples produced several excitations such as CH, OH combination bands, CH, OH, NH overtones (first, second), and CH overtone (third) (Osborne 2006). NIR reflectance intensities were influenced by the presence of various sample contents and hence, peaks and troughs were formed at specific wavelengths. Wavelengths related to the peaks and troughs in the slopes plot of the normalized NIR mean reflectance intensities were similar to the results obtained by Wang et al. (1999), Delwiche and Massie (1996), and Murray and Williams (1987). They further reported that wavelengths at 960, 1420, 1470, and 1500 nm were related to water and protein, wavelengths at 1200, 1230, 1310, 1360, 1610, and 1700 nm were related to carbohydrate, wavelengths at 960 nm, 1060 nm, 1330 nm, 1390 nm, 1480 nm, and 1680 nm were related to kernel hardness, and wavelength at 1390 nm was related to oil content in wheat.

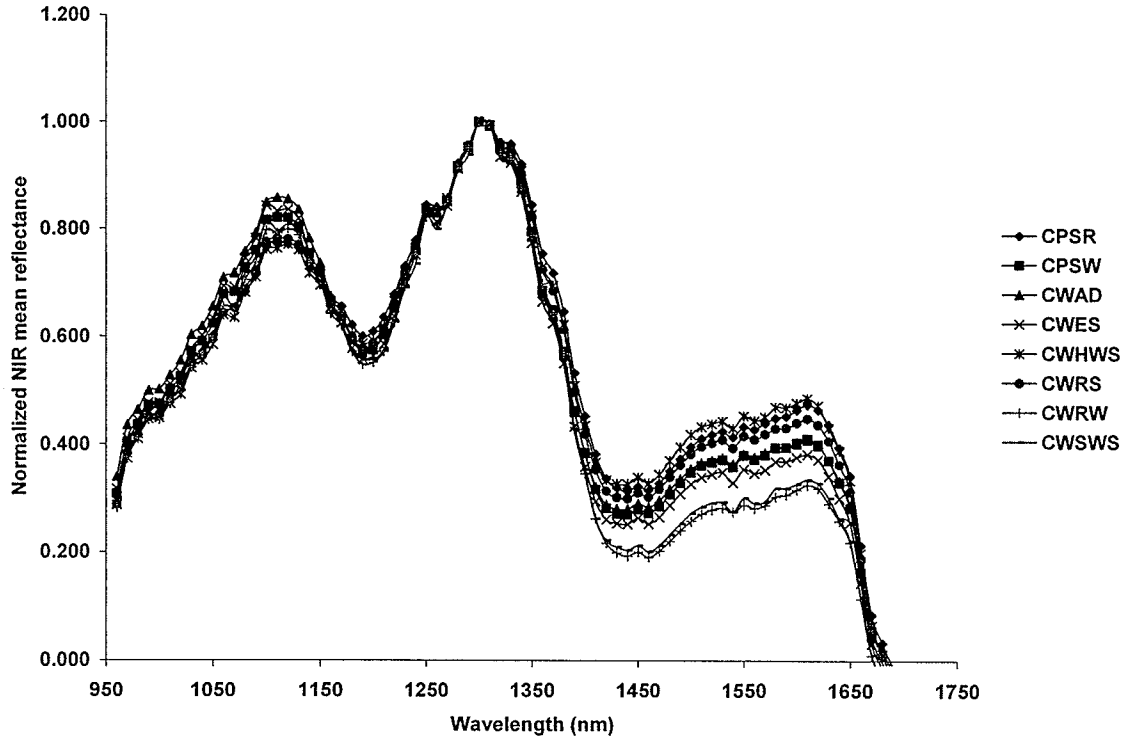

Fig. 4.1 NIR reflectance spectra of wheat classes.

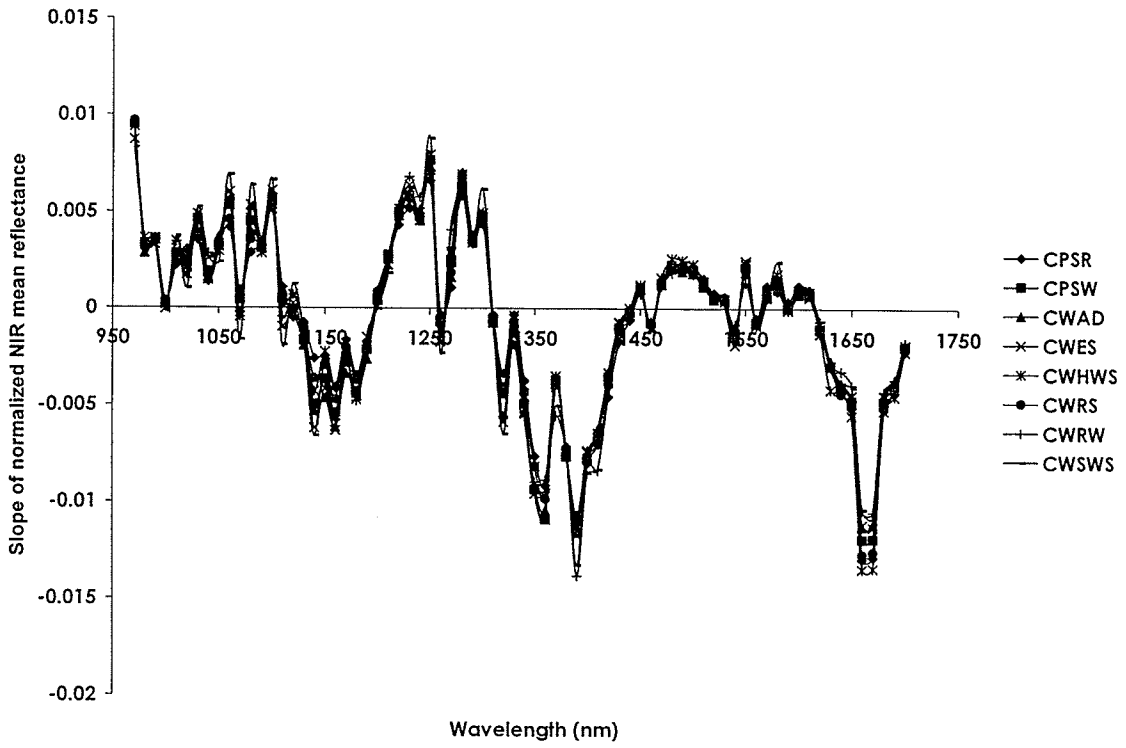

Fig. 4.2 Slopes of NIR reflectance spectra of wheat classes.

4.1.2 Classification accuracies using LDA, QDA, and ANN

The normalized NIR mean reflectance intensities at each wavelength slice of NIR hyperspectral images of bulk wheat samples were used as input for differentiating the wheat classes. Classification accuracies of wheat classes at 11% moisture level using LDA and QDA (n = 300) with leave-one-out cross validation are shown in Fig. 4.3. The classification accuracies were 100% for CPSR, CWES, CWHWS, CWRS, CWRW, and CWSWS wheat using LDA with leave-one-out cross validation. For CPSW and CWAD wheat, classification accuracies were 98.67%. About 1% of CPSW wheat was misclassified as CWRS wheat.

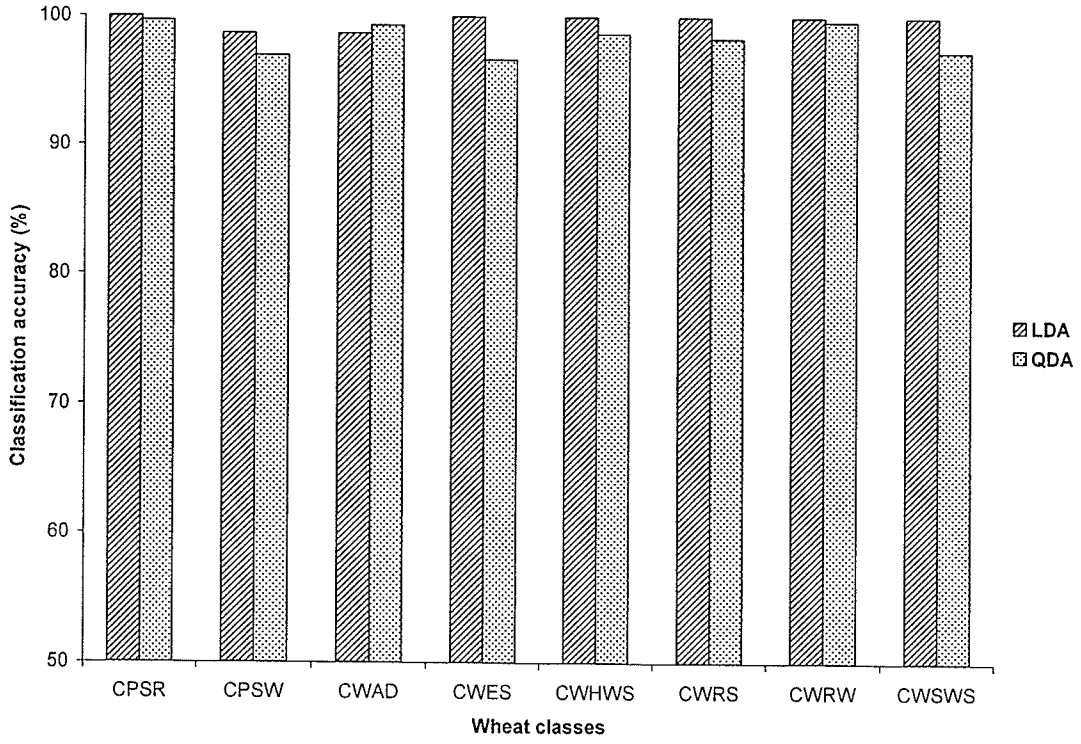

Fig. 4.3 Classification accuracies of wheat classes using LDA and QDA with leave-one-out cross validation (n = 300).

Classification accuracy was 99.67% for CPSR wheat using QDA with leave-one-out cross validation. Classification accuracies of 99.67, 99.33, 98.33, 98.67, 97.00, 96.67, and 97.33% were obtained for CWRW, CWAD, CWRS, CWHWS, CPSW, CWES, and CWSWS wheat, respectively. Classification accuracy of CWES wheat was < 97%. Nearly 3% samples of CWES wheat were misclassified as CPSW wheat. In total, nearly 3% of samples of CWSWS wheat class were identified as all other classes of wheat except CWES, CWHWS, and CWRS wheat class.

Table 4.1 shows the top ten wavelengths determined using STEPDISC procedure and ranked by their level of contribution to classification. The 1580 nm wavelength was identified as the most significant wavelength with ASCC value of 0.12 and the 1160 nm

wavelength was ranked 10 with ASCC value of 0.55. Normalized NIR mean reflectance spectra of wheat classes had the highest peak at 1310 nm wavelength. It could be seen from the slopes plot of normalized NIR mean reflectance spectra that the top ten wavelengths were at and around the peaks or troughs of the plot. The level of contribution of the top ten wavelengths in wheat class identification was high.

Table 4.1 Top ten wavelengths of NIR hyperspectral images of wheat classes using STEPDISC procedure based on their contribution to classification.

No.	Wavelength (nm)	Partial R^2	ASCC
1	1580	0.88	0.12
2	1380	0.89	0.24
3	960	0.74	0.31
4	1490	0.55	0.37
5	1200	0.50	0.43
6	1700	0.37	0.45
7	1280	0.34	0.48
8	1100	0.28	0.50
9	1610	0.28	0.51
10	1160	0.27	0.55

Osborne (2006) observed that wavelengths responsible for peaks and troughs at wavelength regions of 960 – 1100 nm and 1420 – 1600 nm were related to NH, OH overtones (second) and NH, OH overtones (first), respectively. Further study showed that these wavelengths were related mostly to water and protein contents of the sample. Also, wavelengths at 1100 – 1300 nm and 1600 – 1800 nm were related to CH first and second overtones, respectively. Armstrong (2006) observed that 950 – 975 nm and 1400 – 1450 nm wavelength regions were associated to water absorption from the moisture content prediction studies in soybean.

Mohan et al. (2005) reported mean classification accuracies of 89.1% and 99.1% in classifying seven cereal grains using a linear parametric classifier with the top two and

five reflectance features in visible wavelength regions. Delwiche and Graybosch (2002) reported classification accuracies of 42 – 71% in identifying waxy wheat using linear discriminant function with 1 – 10 principal component scores as input. They also reported classification accuracies of 46 – 71% using quadratic discriminant function with 1 – 10 principal component scores as input.

Classification accuracies of wheat classes using 60-30-10 and 70-20-10 ANN models are shown in Figs. 4.4 and 4.5, respectively. R^2, MSE, MAE, and r values are shown in Table 4.2. Training sets of 60-30-10 and 70-20-10 ANN models showed high correlation coefficient of 0.994. For the 60-30-10 model, the validation set of CPSR wheat had a high classification accuracy of 100%. For the 70-20-10 model, the classification accuracies were 99.5%, 100%, and 100% for training, test, and validation sets of CWSWS wheat, respectively. In the 70-20-10 ANN model, all wheat classes had good classification accuracies of 89 – 100%. The lowest classification accuracy of 80% was obtained for the validation set of CPSW wheat using the 60-30-10 ANN model. The 70-20-10 ANN model produced a relatively low classification accuracy of 89.26% for the training set of CWRS wheat.

Table 4.2 Details of statistical parameters of 60-30-10 and 70-20-10 ANN models.

Statistical parameters	60-30-10 ANN model			70-20-10 ANN model		
	Training	Test	Validation	Training	Test	Validation
R^2	0.988	0.984	0.978	0.989	0.986	0.989
MSE	0.06	0.08	0.12	0.05	0.07	0.05
MAE	0.14	0.17	0.19	0.131	0.155	0.148
r	0.994	0.992	0.989	0.994	0.993	0.994

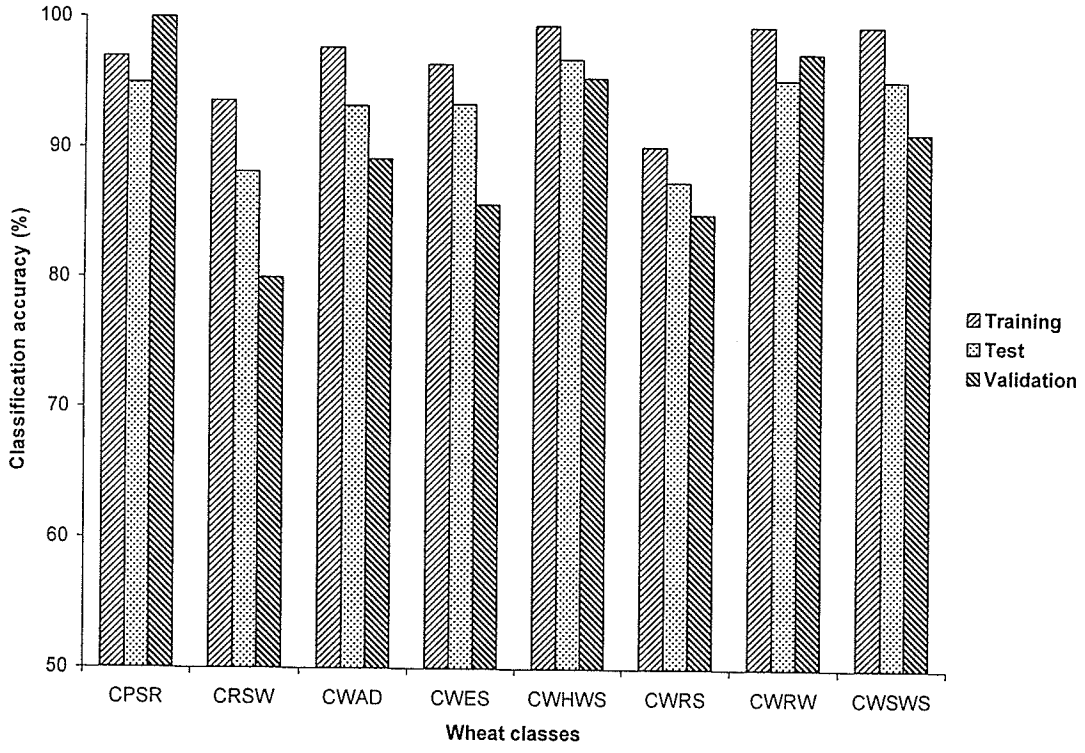

Fig. 4.4 Classification accuracies of wheat classes using the 60-30-10 ANN model (n = 300).

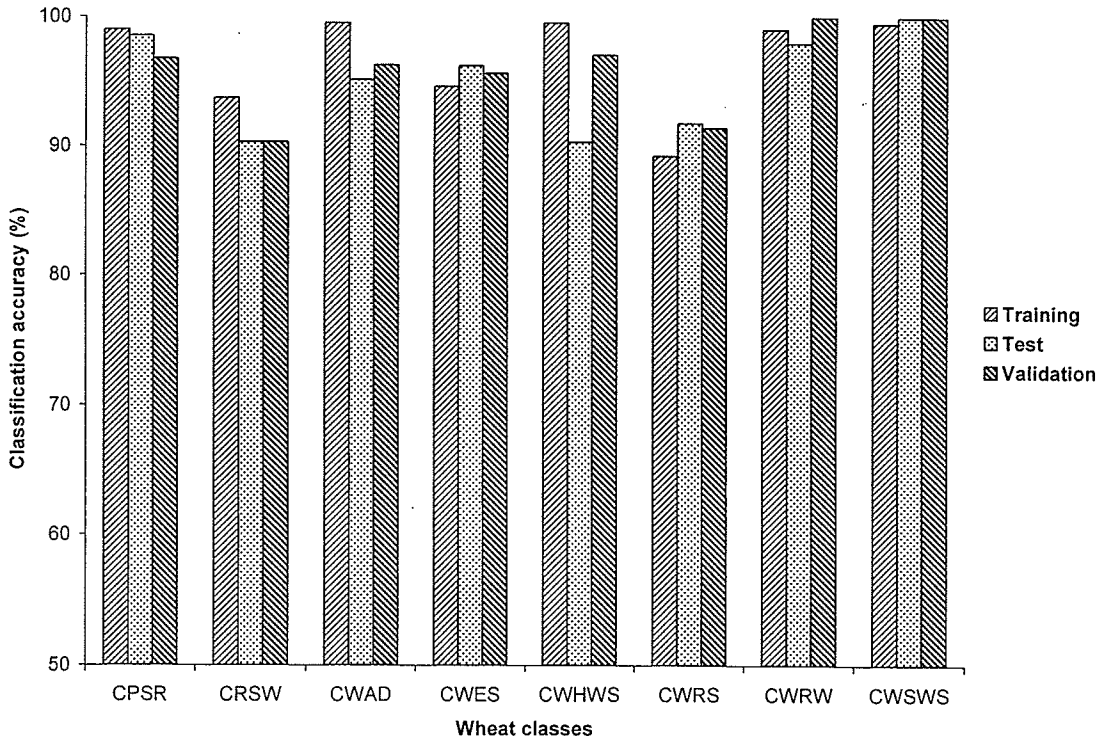

Fig. 4.5 Classification accuracies of wheat classes using the 70-20-10 ANN model (n = 300).

The top 10 wavelengths based on their input strength to wheat class identification using both 60-30-10 and 70-20-10 ANN models are listed in Table 3. Five wavelengths remained the same, but in a different order, for both ANN models. These wavelengths were between 960 and 1300 nm. Osborne (2006) observed that this wavelength region was related to CH second overtone, OH combination band, NH second overtone, OH second overtone, and CH third overtone. A high contribution coefficient of 0.0174 at 1380 nm and a low contribution coefficient of 0.0156 at 1590 nm were attained in the 60-30-10 ANN model. In this training pattern, 1590 nm was the only wavelength that was above the wavelength region of 960 – 1300 nm. The 1360 nm and 1120 nm wavelengths produced high (0.0203) and low (0.0154) contribution coefficients, respectively in the 70-

20-10 ANN model. In this model, the top ten wavelengths were in the wavelength region of 960 – 1480 nm. Classification accuracies of 99% were reported from two class models developed from MLR and PLSR in comparing hard red wheat and hard white wheat using Vis/NIR reflectance properties (Delwiche and Massie 1996). Average classification accuracies of 50 – 89% were reported using ANN with three hidden layer nodes in classifying different types of damages in soybean (Wang et al. 2002). Wang et al. (2004) reported mean classification accuracies of 83 – 95% using ANN in classifying different types of fungal damage in soybean. Classification accuracies were more than 97% for HRS and CWAD in a four layer BPNN in comparing the above mentioned two wheat classes with barley, oats and rye using their morphological features (Paliwal et al., 2001). Mohan et al. (2005) reported a classification accuracy of 92.4% in BPNN using top five slope features of wavelengths in the visible region as input in classifying cereal grains. They also reported classification accuracies of 70 – 90% and 50 – 70% using two and five slope features of wavelengths in NIR region in BPNN, respectively. The statistical linear parametric classifier produced better classification accuracies than BPNN using slope features in the NIR region (Mohan et al. 2005).

In Canada, wheat classes are identified using kernel visual distinguishability (KVD) characteristics of wheat (CGC 2006b). CWRS and CWAD wheat classes are considered as major wheat classes whereas the remaining six wheat classes as minor wheat classes in Canada. Based on Canada western standard committee recommendations, CWRS wheat can have 0.75 – 3.8% of contrasting wheat classes (wheat classes of different color) and 2.3 – 7.5% of other wheat classes in it based on its grade ranging from 1 – 4 (CGC 2007b). In this study, LDA and QDA had higher

classification accuracies of > 96% in differentiating wheat classes than 60-30-10 and 70-20-10 ANN models.

Table 4.3 Top ten wavelengths of NIR hyperspectral images based on their input strength to classification using 60-30-10 and 70-20-10 ANN models in identifying wheat classes.

S. No.	60-30-10 model		70-20-10 model	
	Wavelength (nm)	Contribution coefficient	Wavelength (nm)	Contribution coefficient
1	1380	0.0174	1360	0.0203
2	1490	0.0174	1380	0.0202
3	1360	0.0169	1190	0.0179
4	1480	0.0164	1480	0.0167
5	1190	0.0161	1130	0.0164
6	1340	0.0161	1180	0.0160
7	1370	0.0161	1060	0.0160
8	1580	0.0158	1340	0.0159
9	1090	0.0158	1400	0.0158
10	1590	0.0156	1120	0.0167

4.2. Wheat class differentiation with and without moisture effects (Study: 2)

4.2.1. Wheat class differentiation with moisture effect using statistical classifier

4.2.1.1 Classification accuracies using LDA and QDA (75 input features)

The NIR absorbance values at each wavelength slice of a NIR hyperspectral image of wheat samples were used as input for differentiating the wheat classes. Classification accuracies of wheat classes at various moisture levels (12%, 14%, 16%, 18% and 20%) using LDA and QDA (n = 100 per treatment) with leave-one-out cross validation are shown in Figs. 4.6 and 4.7, respectively.

The classification accuracies were > 94% for all wheat classes except CWES wheat at 20% moisture level and CWHWS wheat at 14% moisture level using LDA with leave-one-out cross validation. For these wheat classes, classification accuracies were only 86%. CWES wheat at 20% moisture level was mainly misclassified with CWRS and CWHWS wheat classes at 20% moisture level. Misclassification between CWHWS wheat at 14% moisture level and CWRS wheat at 12% moisture level was also seen. It showed that wheat classes of high moisture levels were misclassified with other high moisture wheat class or classes. Also, low moisture level wheat was misclassified with other low moisture level wheat. The medium or high protein level wheat classes were misclassified with each other. Misclassifications were not seen among the low protein wheat classes. Except CWES wheat at 20% moisture level, all other moisture levels of CWES wheat attained classification accuracies of ≥ 94% and CWES wheat at 14% moisture level had classification accuracy of 100%. Classification accuracies were ≥ 96% for CWHWS wheat at all moisture levels except at 14% moisture level. CWRS wheat had classification accuracies of ≥ 96% for all moisture levels and moisture levels of 14% and

20% attained a classification accuracy of 100%; and 18% moisture level wheat attained a classification accuracy of 99%. CWRW wheat attained classification accuracies of ≥ 98% for all moisture levels and 14% moisture level had classification accuracy of 100% using LDA. CWSWS wheat had good classification accuracies. CWSWS wheat of 14 – 20% moisture levels had a classification accuracy of 100%. A classification accuracy of 98% was attained by 12% moisture level of CWSWS wheat.

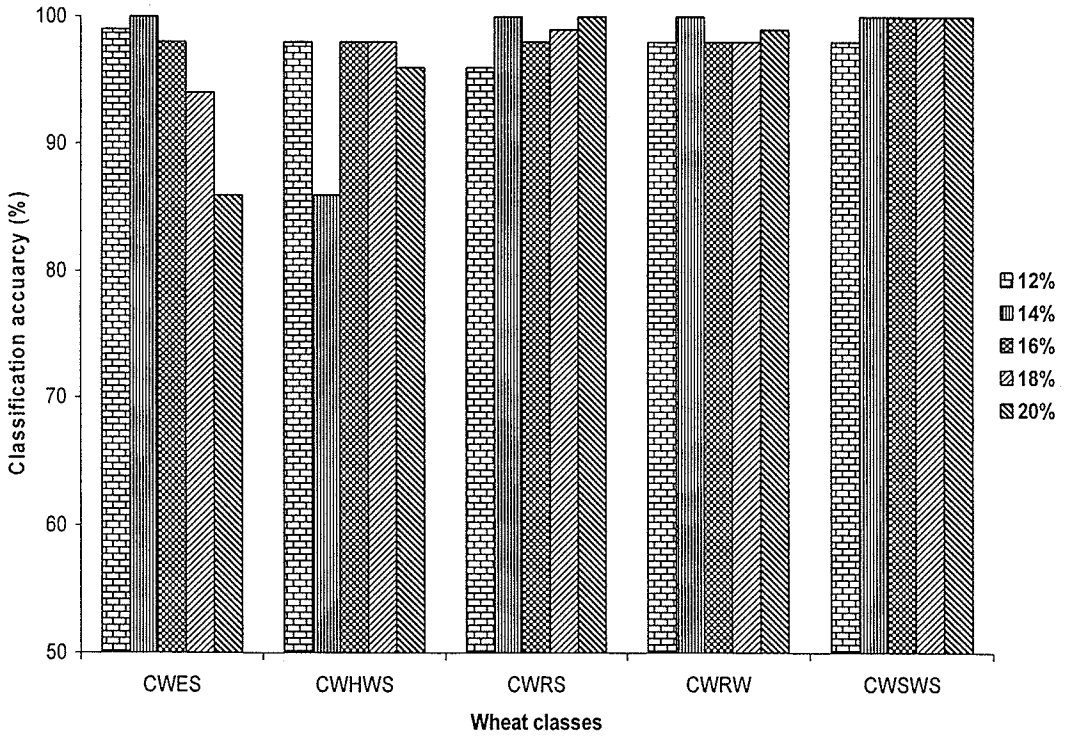

Fig. 4.6 Classification accuracies of wheat classes at various moisture levels using LDA with leave-one-out cross validation (75 input features).

In wheat class identification at various moisture levels, QDA did not perform as well as LDA. Leave-one-out cross validation was used in QDA also. None of the wheat classes attained a classification accuracy of 100% using QDA. But, most of them attained classification accuracies of > 90%. In CWES wheat, classification accuracies were ≥ 98%

for all moisture levels except 18% moisture level. CWES wheat at 18% moisture level had a classification accuracy of only 85%. Misclassification mainly occurred in 18% moisture level of CWES wheat with 18% and 20% moisture levels of CWHWS wheat and 18% moisture level of CWRS wheat. Poor classification accuracies were reported for CWHWS wheat classes at all moisture levels. CWHWS wheat at high and low moisture levels (20% and 12%) had a classification accuracy of only 77%. CWHWS wheat at 12% moisture level was mainly misclassified with other high protein wheat such as CWRS at 12% moisture level. CWHWS wheat at 20% moisture level was misclassified as 18% moisture level of CWES wheat and 20% moisture level of CWRS wheat. It was proven in QDA also, that misclassifications among wheat classes were mainly based on their protein and moisture levels. CWRW wheat had classification accuracies of $\geq$ 90% for all moisture levels except 12% moisture level. CWRW wheat at 12% moisture level had a classification accuracy of 85% and was mainly misclassified with the same wheat class at 14% moisture level. CWSWS wheat at all moisture levels had classification accuracies of $\geq$ 92%.

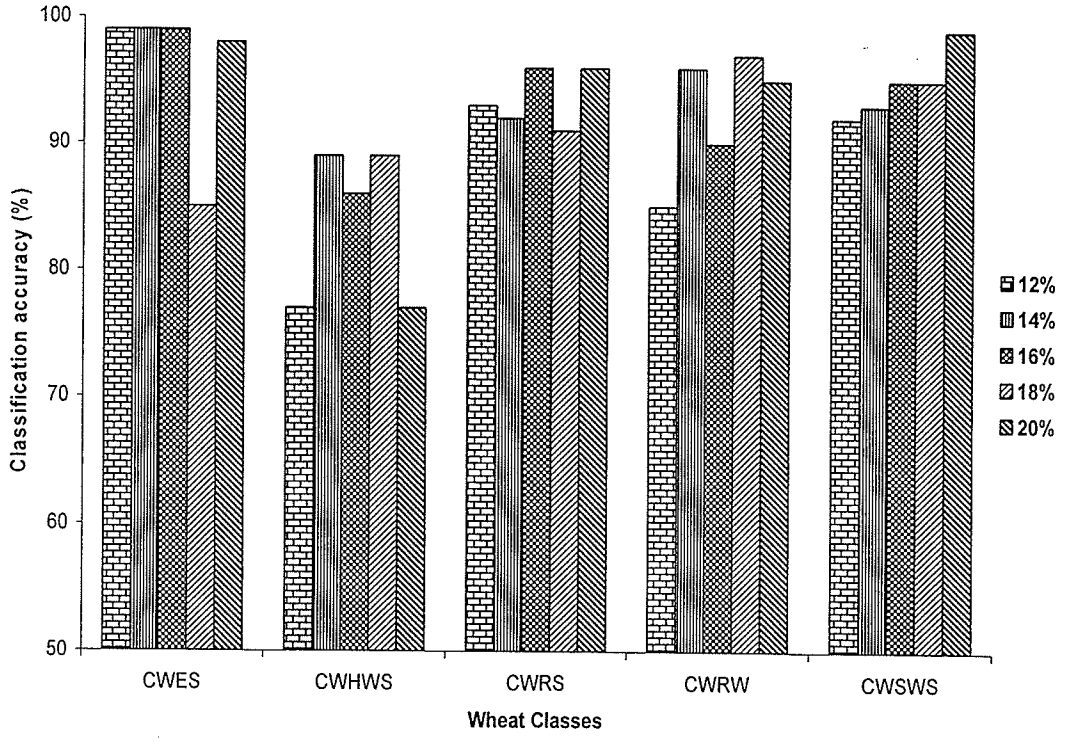

Fig. 4.7 Classification accuracies of wheat classes at various moisture levels using QDA with leave-one-out cross validation (75 input features).

NIR absorbance features at each wavelength had some contribution in wheat class identification at various moisture levels. Using STEPDISC procedure in SAS, the features were ranked based on their contribution to wheat class identification. Feature contributions were explained in terms of partial R^2 and ASCC values. Top ranked input feature was related to 1100 nm wavelength and it had partial R^2 and ASCC values of 0.72 and 0.03, respectively. The last ranked input feature was derived from 1440 nm wavelength and it had partial R^2 and ASCC values of 0.02 and 0.50, respectively. The top ten features with their partial R^2 and ASCC values are shown in Table 4.4. The top ten features had cumulative ASCC value of 0.24. STEPDISC results showed that all input

features should be necessary for wheat class identification purposes. And, reduction in input features led to further reduction of classification accuracies of wheat classes.

Table 4.4 Top ten wavelengths of NIR hyperspectral images based on their contribution to classification using STEPDISC procedure in identifying wheat classes at various moisture levels (75 input features).

No.	Wavelength (nm)	Partial R^2	ASCC
1	1100	0.72	0.03
2	1330	0.80	0.05
3	1580	0.81	0.09
4	1370	0.87	0.12
5	1400	0.78	0.14
6	1210	0.74	0.17
7	960	0.59	0.19
8	1000	0.51	0.21
9	1240	0.49	0.22
10	1700	0.43	0.24

4.2.1.2 Classification accuracies using LDA and QDA (51 input features)

Classification accuracies of wheat classes at various moisture levels using LDA with 51 NIR absorbance features in the wavelength region of 1100 – 1600 nm are given in Fig. 4.8. CWES wheat at moisture levels of 12, 14, and 16% had classification accuracies of ≥ 98% and 14% moisture level of CWES wheat had a classification accuracy of 100%. But, 18% and 20% moisture levels of CWES wheat had poor classification accuracies of 86% and 71%, respectively. CWES wheat at 18% moisture level was mainly misclassified with CWHWS wheat at 18% moisture level. CWES wheat with 20% moisture level was misclassified as 20% moisture level of CWHWS wheat and 20% moisture level of CWRS wheat. Same trend was observed in LDA with 75 input features. Misclassification occurred mainly among the wheat classes of similar moisture and protein levels. CWHWS wheat at all moisture levels had classification accuracies of ≥ 92% and 12% moisture level of CWHWS wheat had a classification accuracy of 100%.

CWRS wheat at all moisture levels also had classification accuracies of ≥ 94% in LDA. CWRS wheat at 14, 16, and 18% moisture levels had classification accuracies of 100, 98, and 98%, respectively. CWRW wheat at all moisture levels also had classification accuracies of > 97% and they were misclassified within the same wheat class of next or previous moisture levels. CWRW wheat at 14 and 20% moisture levels had a classification accuracy of 100% in LDA. CWSWS wheat at 16, 18, and 20% moisture levels had a classification accuracy of 100%. And, the other two moisture levels of CWSWS wheat (12 and 14%) had classification accuracies of 99 and 97%.

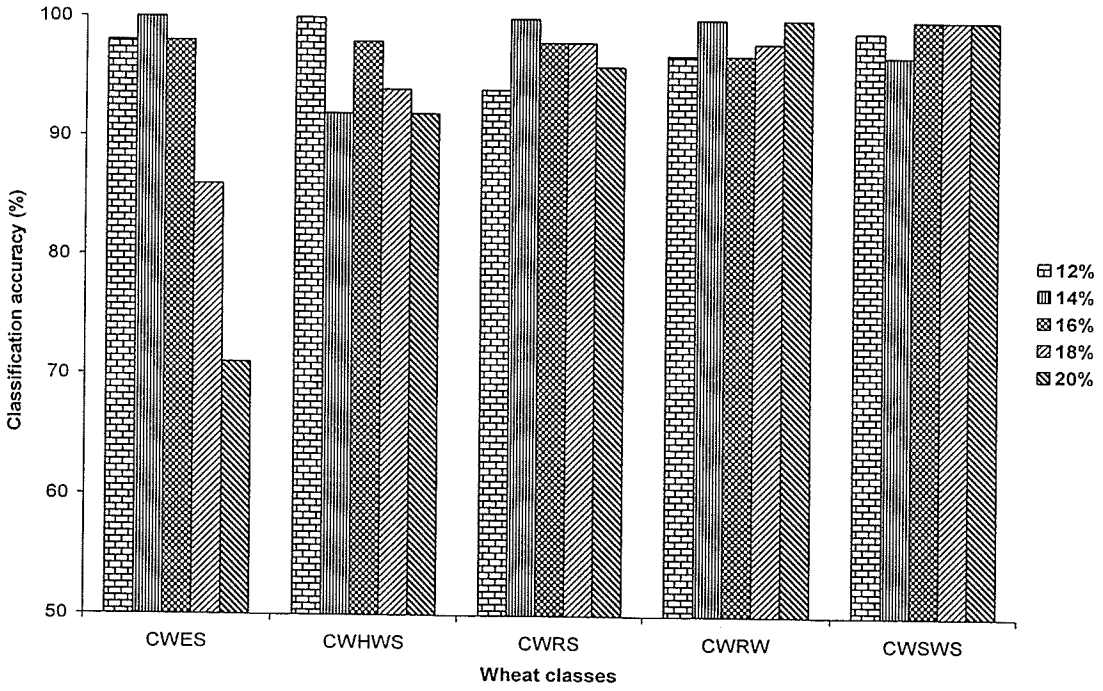

Fig. 4.8 Classification accuracies of wheat classes at various moisture levels using LDA with leave-one-out cross validation (51 input features).

Classification accuracies using QDA for wheat classes with 51 input features are given in Fig. 4.9. Classification accuracies for CWES wheat at all moisture levels were improved in QDA. CWES wheat at 12, 14, and 16% moisture levels had classification

accuracies of 100%. Classification accuracies of 18 and 20% moisture level of CWES wheat were also improved. CWES wheat at 20% moisture level had a classification accuracy of 95%. CWES wheat at 18% moisture level was misclassified as 18 and 20% moisture levels of CWHWS wheat and 18% moisture level of CWRS wheat. Classification accuracies of CWHWS wheat at all moisture levels using QDA were not as good as those of LDA. Poor accuracies were found for 12, 18, and 20% moisture levels of CWHWS wheat. They were mainly misclassified with the wheat classes with similar level of protein and moisture contents. CWRS wheat at 12 – 18% moisture levels had classification accuracies of 98% and 20% moisture level wheat had a classification accuracy of 99%. CWRW wheat at all moisture levels had high classification accuracies of ≥ 92% and 14 – 20% moisture levels had classification accuracies of ≥ 97%. Misclassification occurred with the immediate next moisture level of the same wheat class. CWSWS wheat at all moisture levels had classification accuracies of ≥ 96%; and misclassification occurred mainly with previous or next moisture level of the same wheat class. QDA improved classification accuracies of CWES and CWRS wheat classes.

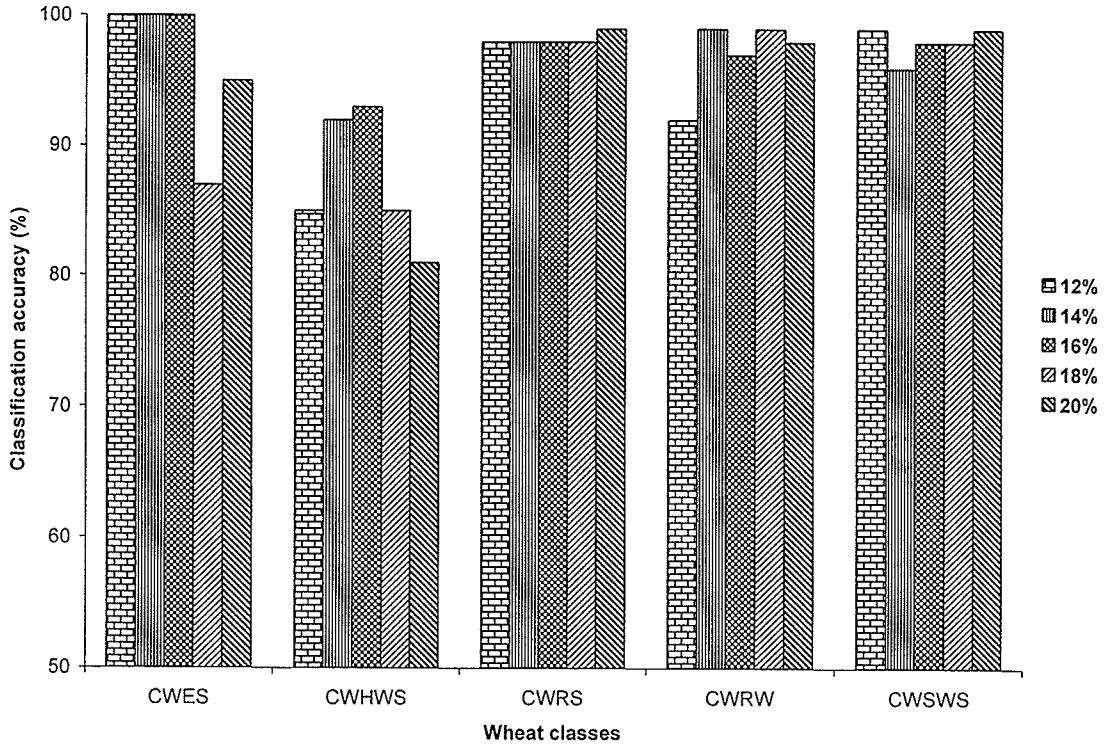

Fig. 4.9 Classification accuracies of wheat classes at various moisture levels using QDA with leave-one-out cross validation (51 input features).

The top ten wavelengths were identified using STEPDISC procedure in SAS. (Table 4.5). Input feature at 1100 nm wavelength was identified as the top feature and at 1440 nm as the last rank feature based on their contribution to classification. These results were in confirmation with the results of STEPDISC with 75 input features. These features were ranked based on partial R^2 and ASCC values. Input feature at 1100 nm had partial R^2 and ASCC values of 0.72 and 0.03, respectively. Input feature at 1440 nm had partial R^2 and ASCC values of 0.06 and 0.41, respectively. The top ten features had a cumulative ASCC value of 0.22. When top ten features were only used for wheat class identification, the classification accuracies were drastically reduced. It showed that all

features should be used for attaining good results in wheat class identification at various moisture levels. Out of 10 top features of the 51 feature model, seven were the same with those of the 75 feature model.

Table 4.5 Top ten wavelengths of NIR hyperspectral images based on their contribution to classification using STEPDISC procedure in identifying wheat classes at various moisture levels (51 input features).

No.	Wavelength (nm)	Partial R^2	ASCC
1	1100	0.72	0.03
2	1330	0.80	0.05
3	1580	0.81	0.09
4	1370	0.87	0.12
5	1400	0.78	0.14
6	1210	0.74	0.17
7	1190	0.47	0.18
8	1240	0.46	0.20
9	1150	0.46	0.21
10	1180	0.41	0.22

4.2.1.3 Classification accuracies using LDA and QDA (top 7 input features)

Top seven common wavelengths were identified using STEPDISC procedure with 75 and 51 input features. They were 1100, 1330, 1580, 1370, 1400, 1210, and 1240 nm. Wheat class identification was done using LDA and QDA with the NIR absorbance features at the above wavelengths. Classification accuracies of top 7 input features were compared with the classification accuracies of LDA (75 features, 51 features) and QDA (75 features, 51 features). Classification accuracies of wheat classes at various moisture levels using LDA and QDA at 75, 51, and top 7 input features are shown in Figs. 4.10, 4.11, 4.12, 4.13, and 4.14, respectively. The classification accuracies of LDA with the top 7 wavelength features were the same or below to the classification accuracies of LDA with 75 and 51 input features. Classification accuracies of QDA with top 7 wavelength

features had more fluctuations. Mostly, they had three different trends: higher, equal, or lower classification accuracies than QDA with 75 or 51 input features.

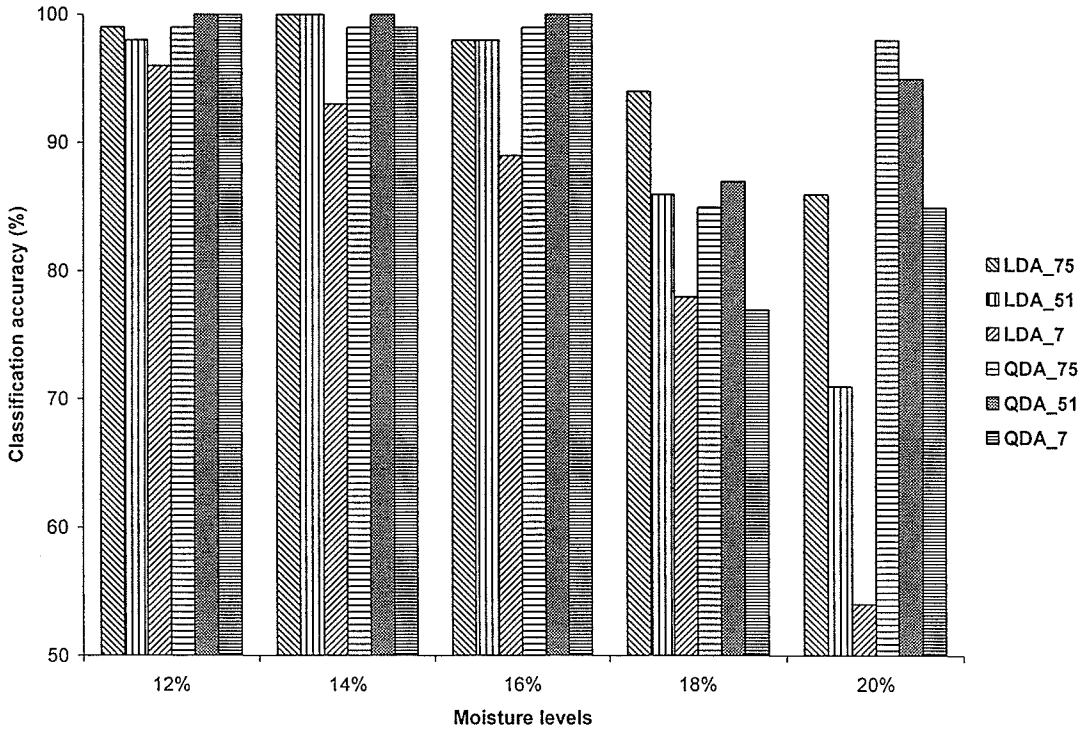

Fig. 4.10 Classification accuracies of CWES wheat a various moisture levels using LDA and QDA at 75, 51, and top 7 input features.

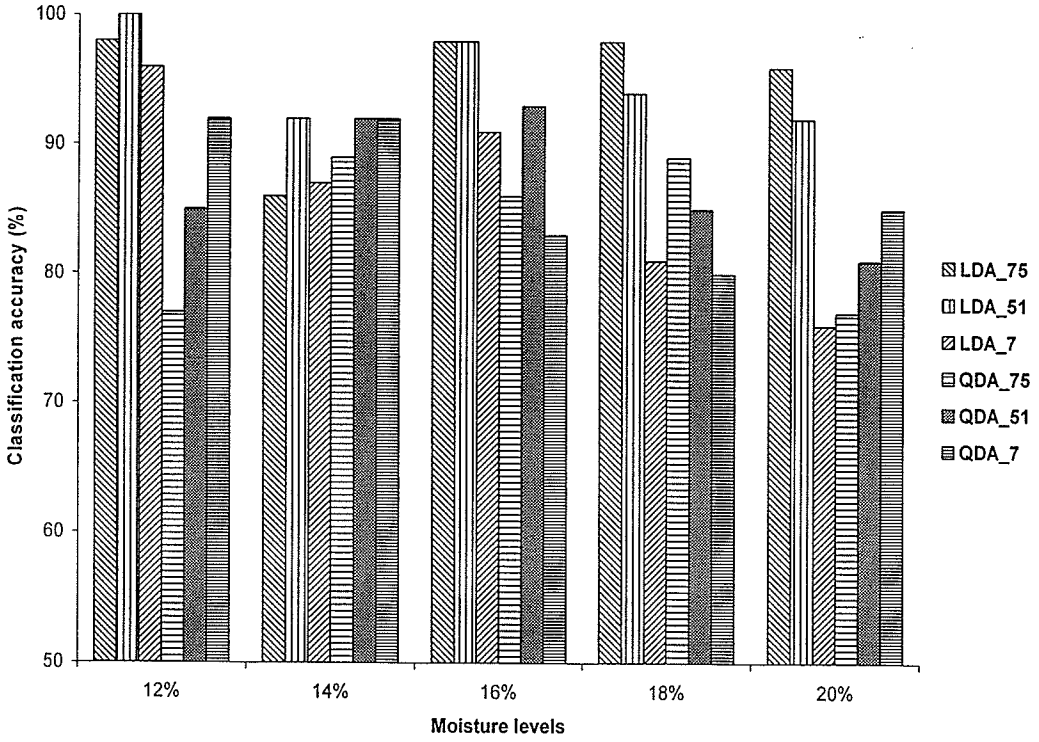

Fig. 4.11 Classification accuracies of CWHWS wheat at various moisture levels using LDA and QDA at 75, 51, and top 7 input features.

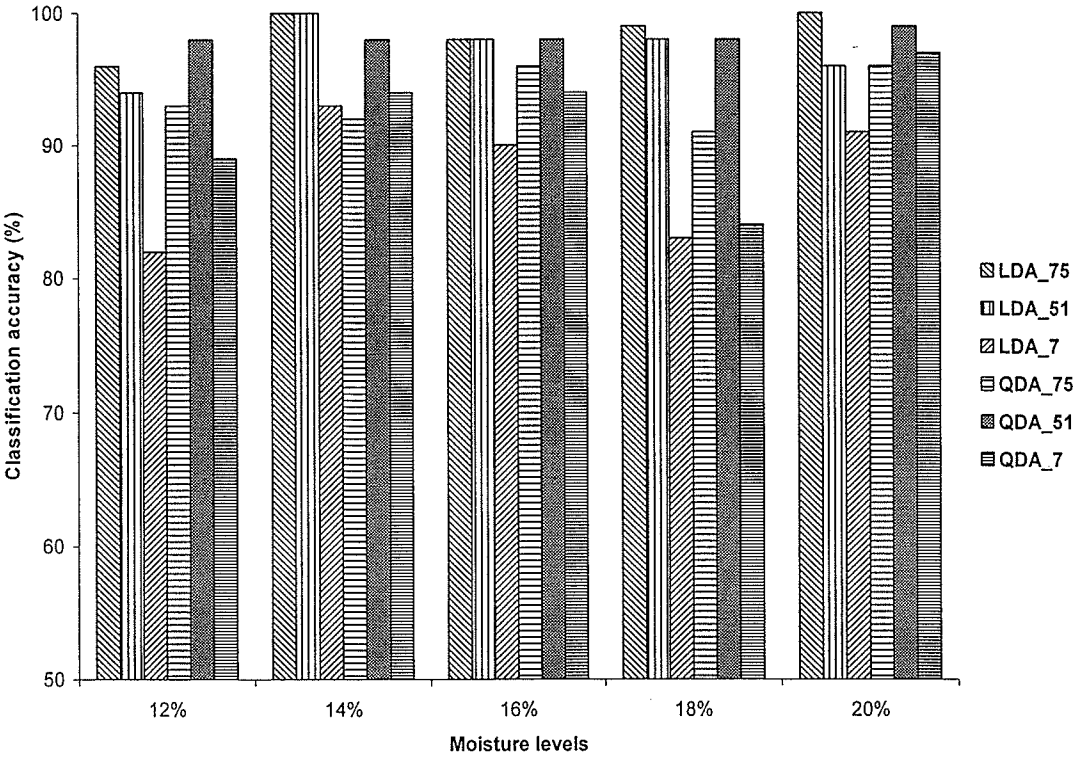

Fig. 4.12 Classification accuracies of CWRS wheat at various moisture levels using LDA and QDA at 75, 51, and top 7 input features.

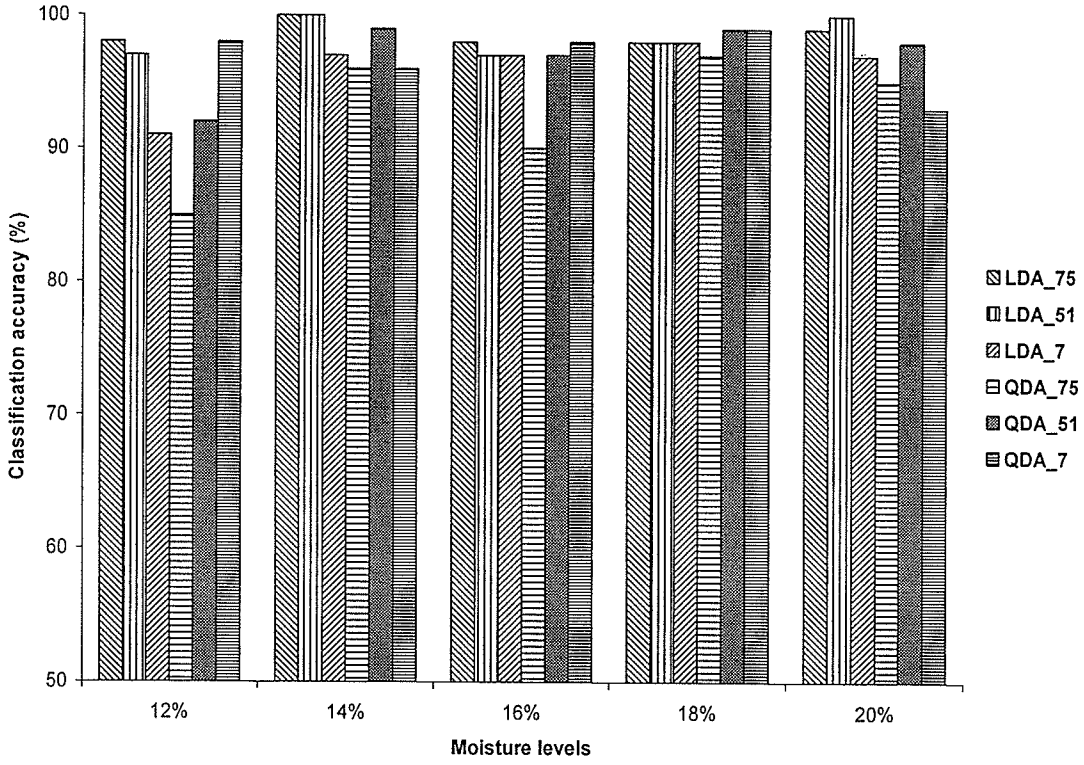

Fig. 4.13 Classification accuracies of CWRW wheat at various moisture levels using LDA and QDA at 75, 51, and top 7 input features.

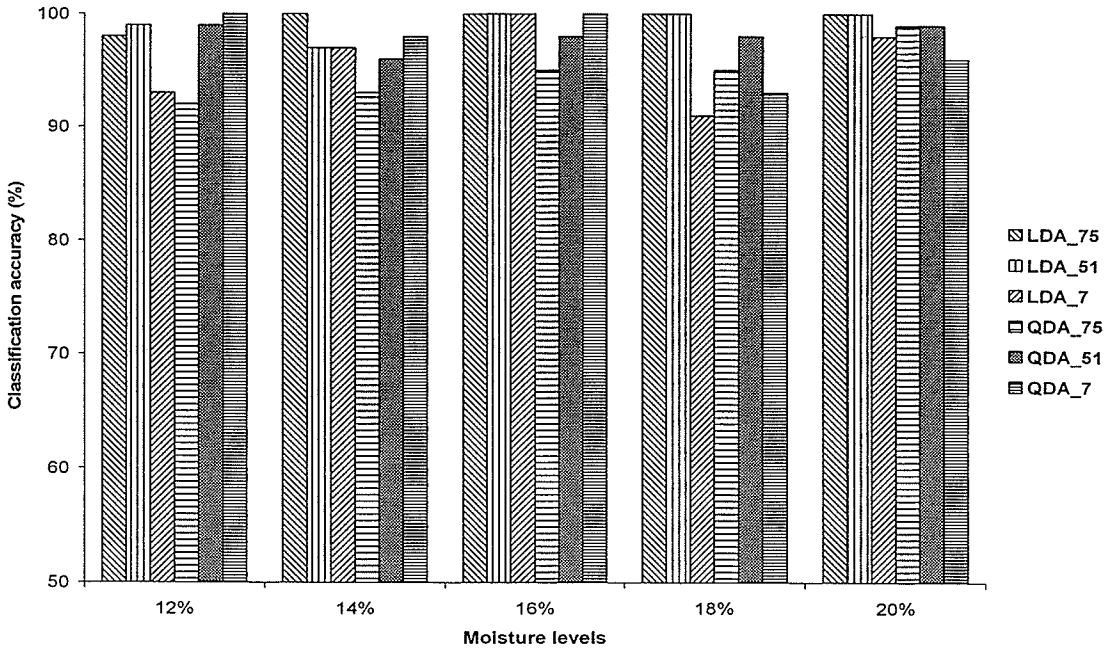

Fig. 4.14 Classification accuracies of CWSWS wheat at various moisture levels using LDA and QDA at 75, 51, and top 7 input features.

4.2.2. Wheat class differentiation with and without moisture effects using ANN (75 input features)

Table 4.6 shows the top ten wavelengths of NIR hyperspectral images based on their input strength to classification using a 60-30-10 ANN model in identifying the wheat classes without moisture effect. Wheat classes at various moisture levels were identified using ANN. Average minimum error for the training and test sets were 0.0022 and 0.0033, respectively. Input feature related to 1100 nm wavelength was at the 1st rank and 1080 nm at the 10th rank. Input strength of 1st rank and 10th rank features were 0.019 and 0.0159, respectively.

Table 4.6 Top ten wavelengths of NIR hyperspectral images based on their input strength to classification using the 60-30-10 ANN model in identifying wheat classes without moisture effect (75 input features).

No.	Wavelength (nm)	Input strength
1	1100	0.0190
2	1190	0.0183
3	1360	0.0168
4	1500	0.0166
5	1180	0.0166
6	1120	0.0165
7	1200	0.0165
8	1130	0.0163
9	1480	0.0161
10	1080	0.0159

Details of statistical parameters of 60-30-10 ANN model for wheat classes without moisture effect are shown in Table 4.7. In this model, R^2 values of the training and test sets were 0.9731 and 0.9604, respectively. The 60-30-10 ANN model gave a minimum MSE value of 0.053 and a maximum MSE value of 0.081 for the training and test sets, respectively. A minimum MAE and a maximum MAE values were attained by the training and test sets of the 60-30-10 ANN model. A maximum correlation (r =

0.9865) was observed for the training set of the 60-30-10 ANN model. The test set of the ANN model attained a minimum r value of 0.9800.

Table 4.7 Details of statistical parameters of the 60-30-10 ANN model for wheat classes without moisture effects (75 input features).

Parameters	Training	Test	Validation
R^2	0.9731	0.9604	0.9615
MSE	0.053	0.081	0.080
MAE	0.150	0.188	0.190
R	0.9865	0.9800	0.9807

In the 60-30-10 ANN model, the training set of CWHWS wheat gave a classification accuracy of 94.22% and the test set of CWHWS wheat gave a classification accuracy of 90% (Table 4.8). For CWES wheat, the training and test sets had classification accuracies of 92% and the validation set gave a classification accuracy of only 87.5%. The training set of CWRS wheat gave a classification accuracy of > 90%. The test and validation sets of CWRS wheat had classification accuracies of 82% and 76%, respectively. CWRW wheat gave good classification accuracies of > 92% for the training, test, and validation sets. The training set of CWRW wheat had a classification accuracy of 95.06%. CWSWS wheat attained a classification accuracy of 100% for the training, test, and validation sets.

Table 4.8 Classification accuracies of wheat classes without moisture effect using the 60-30-10 ANN model (75 input features).

Wheat class	Training (%)	Test (%)	Validation (%)
CWHWS	94.22	90	90.48
CWES	92.83	92.05	87.5
CWRS	91.16	82.48	76.09
CWRW	95.06	92.30	92.5
CWSWS	100	100	100

Wheat classes were also identified with the moisture effect using ANN. For ANN model development, 75 NIR absorbance features in the wavelength region of 960 – 1700 nm were used. The top ten wavelengths of NIR hyperspectral images based on their input strength to classification using the 60-30-10 ANN model in identifying wheat classes with moisture effect are given in Table 4.9.

Table 4.9 Top ten wavelengths of NIR hyperspectral images based on their input strength to classification using the 60-30-10 ANN model in identifying wheat classes with moisture effect (75 input features).

No.	Wavelength (nm)	Input strength
1	1100	0.0187
2	1190	0.0183
3	1120	0.0171
4	1500	0.0167
5	1360	0.0165
6	1180	0.0164
7	1310	0.0163
8	1080	0.0162
9	1130	0.0161
10	1480	0.0160

Details of statistical parameters of the 60-30-10 ANN model in identifying wheat classes with moisture effect are shown in Table 4.10. The training set gave a maximum R^2 value of 0.972 whereas the test set gave a minimum R^2 value of 0.954 for the 60-30-10 ANN model. The correlation results of the training set showed that the training set of the 60-30-10 ANN model would give minimum MSE and MAE values. The test set had a maximum MSE value. The training set gave a maximum r value of 0.986 and the test and validation sets had approximately the same r values.

Table 4.10 Details of statistical parameters of the 60-30-10 ANN model in identifying wheat classes with moisture effect (75 input features).

Parameters	Training	Test	Validation
R^2	0.972	0.954	0.958
MSE	1.40	2.39	2.20
MAE	0.84	1.08	1.09
r	0.986	0.977	0.978

Classification accuracies of wheat classes with moisture effect using the 60-30-10 ANN model (75 input features) are shown in Table 4.11. The validation set of CWHWS wheat at 20% moisture level gave a maximum classification accuracy of 71.42% in that class. The test set of CWHWS wheat at 18% moisture level had a minimum classification accuracy of 4.88% in that class. For CWES wheat class, classification accuracies of < 50% were observed for all moisture levels. The validation set of CWES wheat at 12% moisture level gave the highest classification accuracy of 47.05% in that class. The validation set of CWES wheat at 18% moisture level had a classification accuracy of only 9.09%. Poor classification accuracies of < 40% were observed for CWRS wheat class at various moisture levels. CWRS wheat at 14% moisture level gave a maximum classification accuracy of 36.67% and the validation sets CWRS wheat at 18 and 12% moisture levels had classification accuracies of 0%. For CWRW wheat class, classification accuracies of < 50% were observed for all moisture levels. The training set of CWRW wheat at 16% moisture level gave a maximum classification accuracy of 49.18% and the validation set of CWRW wheat at 20% moisture level had a minimum classification accuracy of 16.67% in that class. The classification accuracies of CWSWS wheat were high. The training set of CWSWS wheat at 12% moisture level gave a

maximum classification accuracy of 90.62% and the validation set of CWSWS wheat at 14% moisture level had a minimum classification accuracy of 33.33%.

Table 4.11 Classification accuracies of wheat classes with moisture effect using the 60-30-10 ANN model (75 input features).

Wheat Class	Moisture levels (%)	Training (%)	Test (%)	Validation (%)
CWHWS				
	20	52.30	52.38	71.42
	18	21.27	4.88	8.33
	16	36.00	35.90	45.45
	14	70.69	62.96	40.00
	12	29.82	25.00	27.27
CWES				
	20	24.13	25.81	27.27
	18	17.86	24.24	9.09
	16	35.48	36.67	37.50
	14	43.33	25.81	33.33
	12	45.61	34.61	47.05
CWRS				
	20	37.88	12.00	11.11
	18	32.35	16.00	0.00
	16	24.14	15.38	6.25
	14	36.67	25.00	16.67
	12	26.15	18.18	0.00
CWRW				
	20	37.29	25.71	16.67
	18	48.48	40.74	42.86
	16	49.18	34.375	42.86
	14	39.21	36.84	36.36
	12	31.34	29.17	33.33
CWSWS				
	20	56.67	38.46	57.14
	18	62.5	50.00	70.00
	16	57.37	41.93	62.5
	14	58.82	65.38	33.33
	12	90.62	79.31	71.43

4.2.3. Wheat class differentiation with and without moisture effects using ANN (51 input features)

Top ten wavelengths of NIR hyperspectral images based on their input strength to classification using the 60-30-10 ANN model for wheat classes without moisture effect are given in Table 4.12. NIR absorbance features were ranked based on their input strength to classification. Maximum input strength was 0.0253 for the input feature at 1100 nm. Input feature at 1520 nm was ranked at tenth position with an input strength of 0.0225.

Table 4.12 Top ten wavelengths of NIR hyperspectral images based on their input strength to classification using the 60-30-10 ANN model in identifying wheat classes without moisture effect (51 input features).

No.	Wavelength (nm)	Input strength
1	1100	0.0253
2	1300	0.0250
3	1190	0.0247
4	1270	0.0242
5	1360	0.0236
6	1120	0.0236
7	1340	0.0233
8	1170	0.0227
9	1290	0.0226
10	1520	0.0225

Details of statistical parameters of the 60-30-10 ANN model for the wheat classes without moisture effect are shown in Table 4.13. The 60-30-10 ANN model gave a maximum R^2 value of 0.9604 and a minimum R^2 value of 0.9480 for the training and test sets, respectively. Minimum MSE and MAE values were obtained by the training set of the 60-30-10 ANN model. The 60-30-10 ANN model gave a maximum MSE and MAE values for the test set. The value of r was maximum (= 0.9800) for the training set and minimum (= 0.9737) for the test set of the 60-30-10 ANN model.

Table 4.13 Details of statistical parameters of the 60-30-10 ANN model in identifying wheat classes without moisture effect (51 input features).

Parameters	Training	Test	Validation
R^2	0.9604	0.9480	0.9509
MSE	0.078	0.106	0.102
MAE	0.195	0.231	0.223
r	0.9800	0.9737	0.9754

Classification accuracies of wheat classes without moisture effect using the 60-30-10 ANN model are shown in Table 4.14. Classification accuracies of wheat classes using the 60-30-10 ANN model with 51 input features were less than that of the 60-30-10 ANN model with 75 input features. CWHWS, CWES, and CWRS wheat classes gave classification accuracies of 85 – 90% except for the validation set of CWRS wheat. The validation set of CWRS wheat had a classification accuracy of 76.09% only. CWRW wheat gave classification accuracies of > 90% except for the test set. The test set of CWRW wheat had a classification accuracy of 87.18%. Classification accuracies for the training, test, and validation sets of CWSWS wheat were > 99%.

Table 4.14 Classification accuracies of wheat classes without moisture effect using the 60-30-10 ANN model (51 input features).

Wheat Class	Training (%)	Test (%)	Validation (%)
CWHWS	86.64	80.63	87.30
CWES	89.08	89.40	82.14
CWRS	88.33	85.40	76.09
CWRW	92.11	87.18	95.00
CWSWS	99.68	99.32	100.00

Top ten wavelengths using the 60-30-10 ANN model to identify wheat classes with moisture effect are shown in Table 4.15. Input feature at 1100 nm wavelength was ranked at the first position with an input strength of 0.0265. Input feature at 1120 nm was ranked at tenth position with an input strength of 0.0220.

Table 4.15 Top ten wavelengths of NIR hyperspectral images based on their input strength to classification using the 60-30-10 ANN model in identifying wheat classes with moisture effect (51 input features).

No.	Wavelength (nm)	Input strength
1	1100	0.0265
2	1360	0.0264
3	1580	0.0246
4	1270	0.0243
5	1190	0.0241
6	1300	0.0239
7	1340	0.0232
8	1170	0.0226
9	1520	0.0220
10	1120	0.0220

Details of statistical parameters of the 60-30-10 ANN model for identifying the wheat classes with moisture effect are shown in Table 4.16. The 60-30-10 ANN model gave a maximum R^2 value of 0.9614 and a minimum R^2 value of 0.9468 for the training and test sets, respectively. MSE and MAE values were minimum for the training set and maximum for the test set of the 60-30-10 ANN model. The training and test sets had r values of 0.9806 and 0.9732, respectively.

Table 4.16 Details of statistical parameters of the 60-30-10 ANN model in identifying wheat classes with moisture effect (51 input features).

Parameters	Training	Test	Validation
R^2	0.9614	0.9468	0.9514
MSE	1.980	2.801	2.546
MAE	1.056	1.262	1.204
r	0.9806	0.9732	0.9757

Classification accuracies of wheat classes with moisture effect using the 60-30-10 ANN model are shown in Table 4.17. Classification accuracies of CWHWS wheat class were ≤ 50%. CWHWS wheat at 18% moisture level gave poor classification accuracies of 2.13%, 0%, and 0% for the training, test, and validation sets, respectively. CWES

wheat class had classification accuracies of ≤ 40% for the training, test, and validation sets of various moisture levels. CWES wheat at 18% moisture level gave poor classification accuracies of 8.93%, 9.09%, and 18.18% for training, test, and validation sets. CWRS wheat had classification accuracies of ≤ 50% for the training, test, and validation sets. CWRS wheat at 20% moisture level gave classification accuracies of 16.67%, 12.00%, and 11.11% for training, test, and validation sets, respectively. Classification accuracies of 11 – 64% were observed for CWRW wheat class. For CWSWS wheat, classification accuracies of 0 – 86% were obtained. The validation set of CWSWS wheat at 14% moisture level gave a classification accuracy of 0%; and the training and test sets of CWSWS wheat at 12% moisture level had classification accuracies of > 80%.

Table 4.17 Classification accuracies of wheat classes with moisture effect using the 60-30-10 ANN model (51 input features).

Wheat class	Moisture levels (%)	Training (%)	Test (%)	Validation (%)
CWHWS				
	20	15.38	19.05	7.14
	18	2.13	0.00	0.00
	16	30.00	23.08	45.45
	14	50.00	37.04	46.67
	12	31.58	25.00	9.09
CWES				
	20	27.59	22.58	27.27
	18	8.93	9.09	18.18
	16	27.42	16.67	25.00
	14	40.00	22.58	22.22
	12	31.58	23.08	17.65
CWRS				
	20	16.67	12.00	11.11
	18	20.59	32.00	14.29
	16	25.86	23.08	25.00
	14	30.00	21.43	33.33
	12	20.00	30.30	50.00
CWRW				
	20	28.81	11.43	33.33
	18	53.03	37.04	42.86
	16	27.87	12.50	42.86
	14	25.49	36.84	63.64
	12	17.91	20.83	22.22
CWSWS				
	20	46.67	15.38	50.00
	18	50.00	67.65	50.00
	16	45.90	29.03	37.50
	14	39.71	46.15	0.00
	12	81.25	86.21	71.43

The top ten wavelengths of statistical classifier and the 60-30-10 ANN model in identifying wheat classes are given in Table 4.18.

Table 4.18 Top ten wavelengths of statistical classifier and the 60-30-10 ANN model in identifying wheat classes with and without moisture effect.

Rank	Statistical classifier		60-30-10 ANN model			
	75 input features	51 input features	75 input features		51 input features	
	With moisture effect	With moisture effect	Without moisture effect	With moisture effect	Without moisture effect	With moisture effect
1	1100	1100	1100	1100	1100	1100
2	1330	1330	1190	1190	1300	1360
3	1580	1580	1360	1120	1190	1580
4	1370	1370	1500	1500	1270	1270
5	1400	1400	1180	1360	1360	1190
6	1210	1210	1120	1180	1120	1300
7	960	1190	1200	1310	1340	1340
8	1000	1240	1130	1080	1170	1170
9	1240	1150	1480	1130	1290	1520
10	1700	1180	1080	1480	1520	1120

The important wavelengths for identifying wheat classes without moisture effect are shown in Fig. 4.15. The 60-30-10 ANN model produced good results in identifying wheat classes without moisture effect.

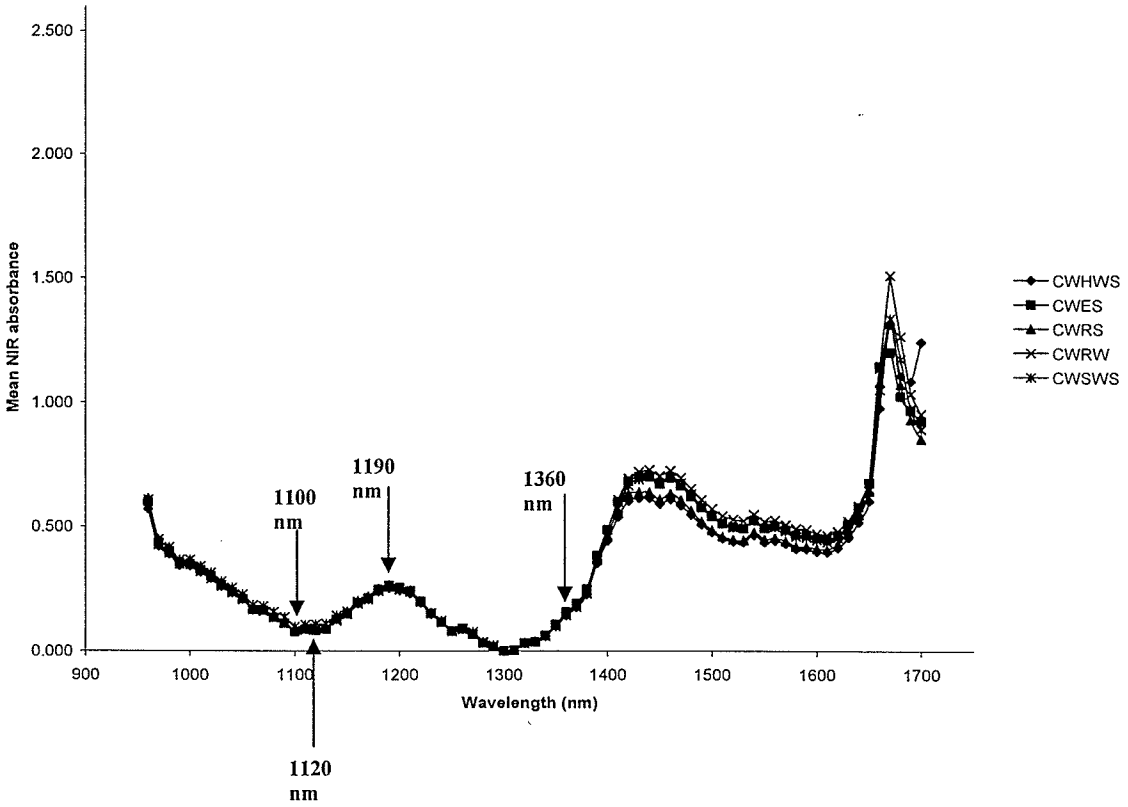

Fig. 4.15 The important wavelengths for identifying wheat classes without moisture effect.

The important wavelengths for identifying wheat classes with moisture effect are shown in Fig. 4.16. Statistical classifier produced good results in identifying wheat classes with moisture effect. The top ten wavelengths based on their contribution to wheat classification were found using STEPDISC procedure in SAS. In Figs. 4.15 and 4.16, a strong peak could be seen at the NIR wavelength of 1440 nm. NIR spectra of wheat classes at various moisture levels were separated from each other at this peak. All wheat classes at various moisture levels had the same trend in the NIR absorbance spectra. Peak at 1440 nm is created by the water molecules of the sample that had different levels of NIR absorption (Osborne et al. 1986). Wang et al. (2004) found that

NIR absorbance peaks at 1440 nm decreased with the decrease in moisture content of ground wheat samples. Armstrong (2006) proved from moisture content prediction study in soybean that the wavelength region of 1400 – 1450 nm was related to the NIR absorption by water molecules of the sample.

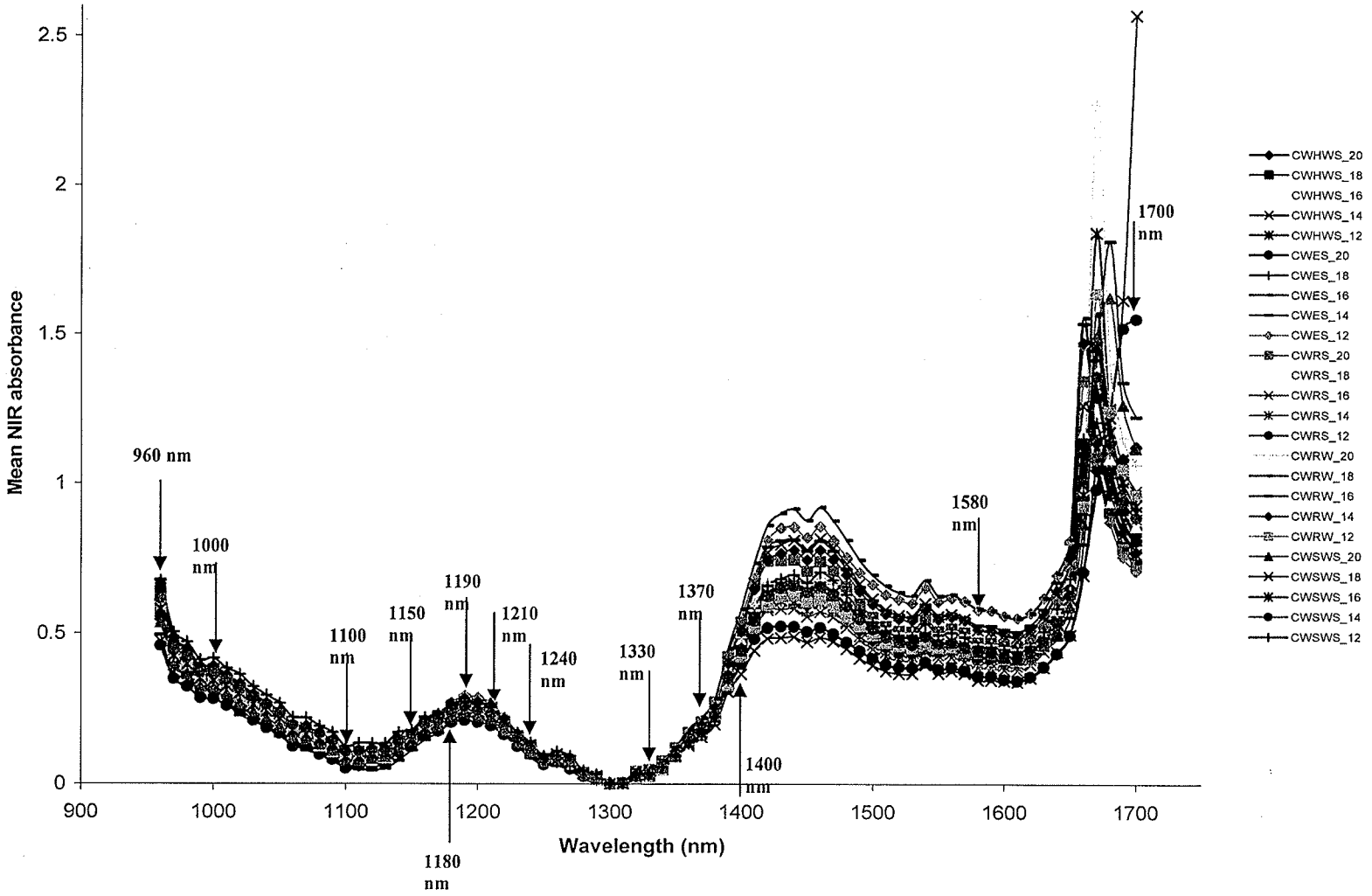

Fig. 4.16 The important wavelengths for identifying wheat classes with moisture effect.

Plots of the first two canonical variables (Can1 vs. Can2) for the wheat classes with and without moisture effects are shown in Figs. 4.17 and 4.18, respectively. Canonical variables were found from the 75 NIR absorbance features of wheat classes using CANDISC procedure in SAS. Canonical variable 1 (Can1) was plotted against canonical variable 2 (Can2) for the wheat classes with and without moisture effects. Can1 had higher correlation with the independent variables than did Can2. From Fig. 4.17, it could be seen that wheat classes at various moisture levels were grouped in different clusters along the axis of Can1 (X axis). High moisture wheat classes were formed into clusters at the front and low moisture wheat classes were formed into clusters at the back of X axis. Wheat classes were grouped in different clusters based on their protein levels along the axis of the Can2 (Y axis). High protein wheat classes were clustered at the bottom whereas low protein wheat classes were clustered at the top of Y axis. Twenty four canonical variables were extracted out of which Can1 and Can2 attained r values of 0.99 and 0.97, respectively.

From Fig. 4.18, it could be seen that wheat classes at various protein levels were grouped in different clusters along the axis of Can1 (X axis). High protein wheat classes were clustered at the front whereas low protein wheat classes were clustered at the back of X axis. Can2 did not help much to explain the variations in the major constituent of wheat without moisture effect. Here, four canonical variables were extracted. Can1 and Can2 attained correlation values of 0.96 and 0.85, respectively. It is proved that NIR absorbance values of hyperspectral imaging could differentiate wheat classes at various moisture and protein levels. A list of wheat classes and their moisture levels for Fig. 4.17 is given in the Appendix 7.2.

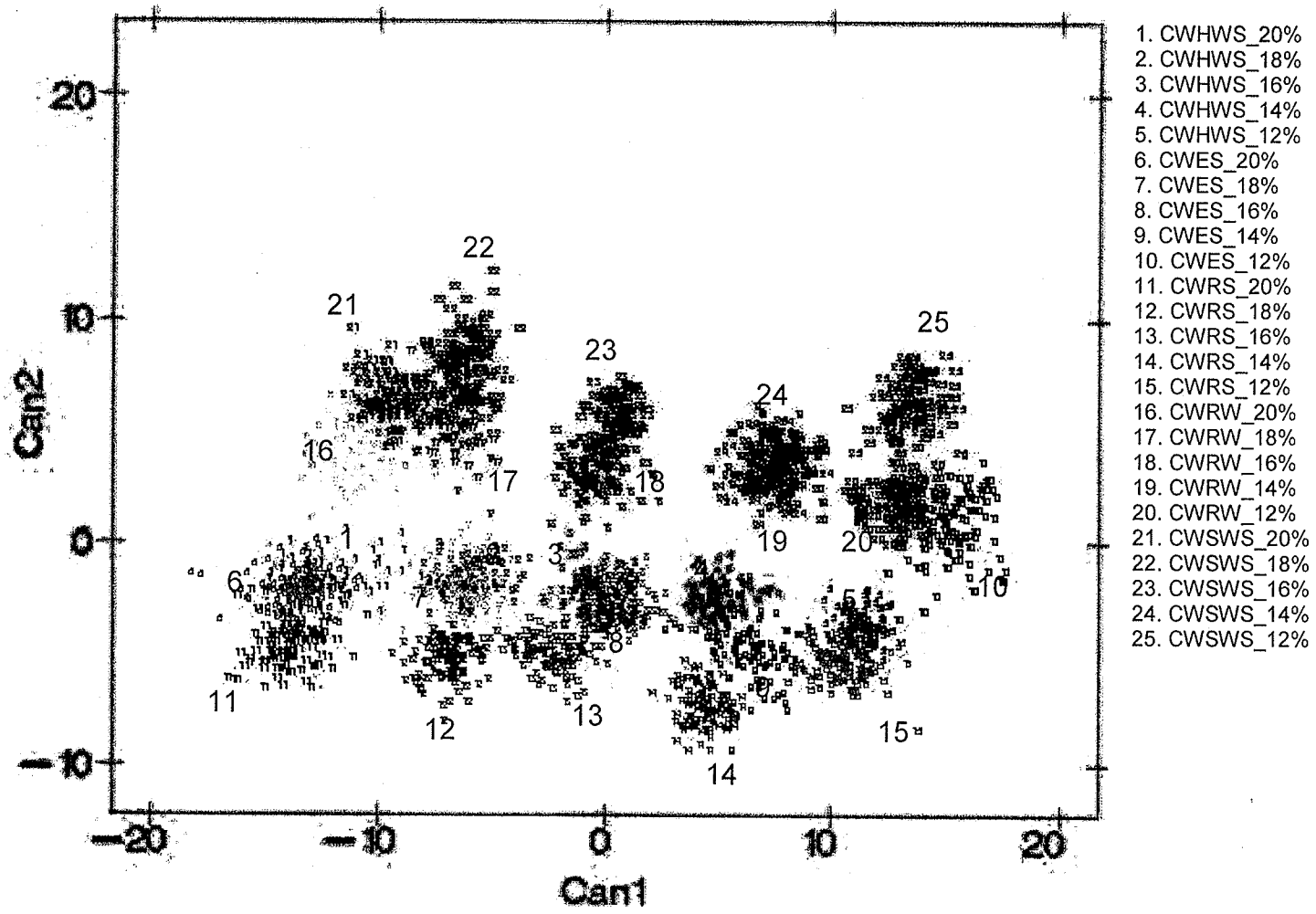

Fig. 4.17 Plot of first two canonical variables for wheat classes with moisture effect.

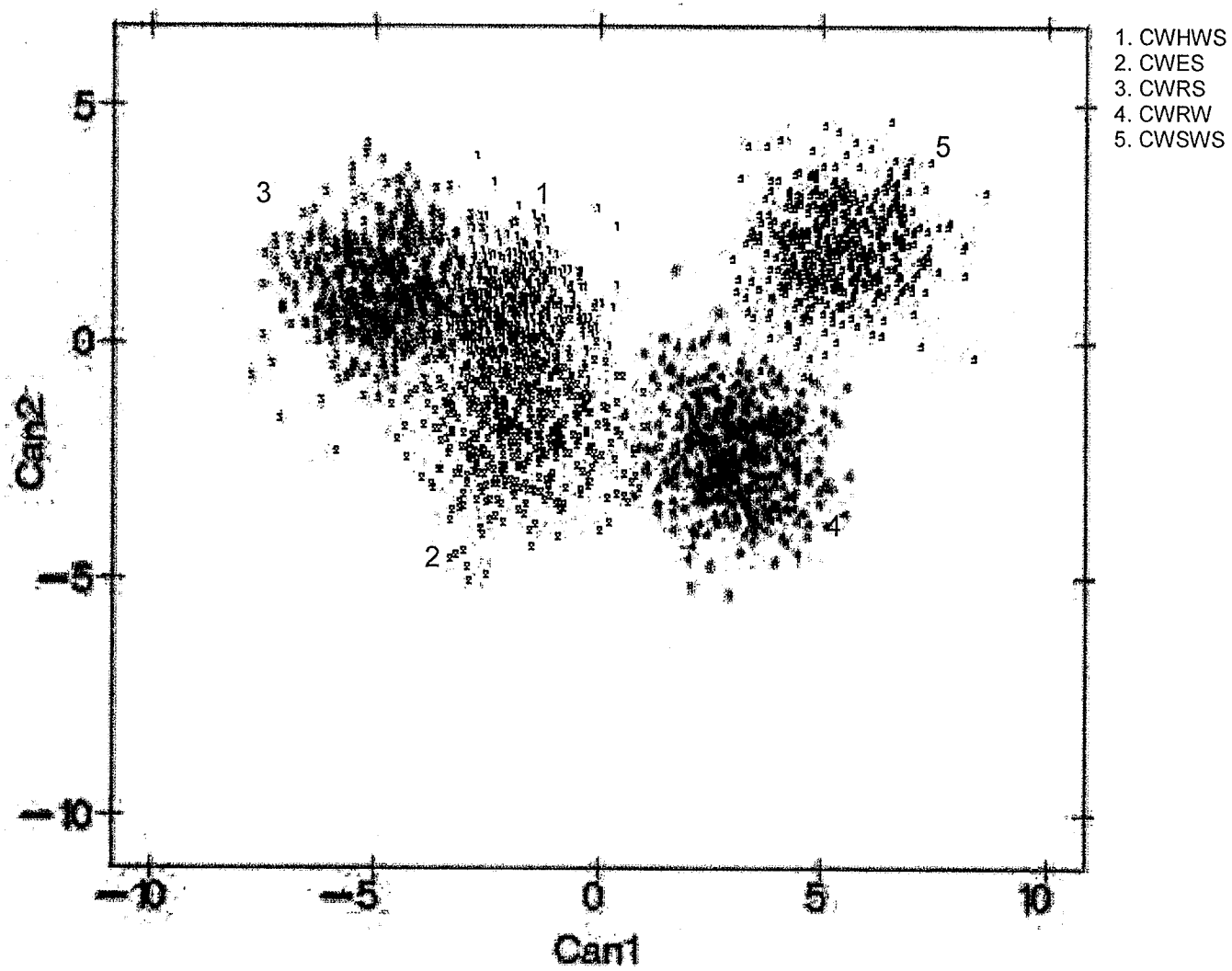

Fig. 4.18 Plot of first two canonical variables for wheat classes without moisture effect.

4.2.4 PCA for wavelength identification in wheat classification

The top ten principal components (PC) were identified using STEPDISC procedure with PC scores as input in SAS. The top ten principal components, their partial R^2 values, and ASCC values in wheat class identification are given in Table 4.19.

Table 4.19 Top ten principal components of NIR hyperspectral images based on their contribution to wheat class identification.

No.	Principal component	Partial R^2	ASCC
1	2	0.89	0.03
2	3	0.84	0.07
3	5	0.90	0.10
4	6	0.77	0.12
5	8	0.77	0.15
6	1	0.76	0.17
7	4	0.75	0.19
8	10	0.63	0.21
9	7	0.53	0.22
10	9	0.49	0.23

Table 4.20 shows the top ten wavelengths of principal components 2, 3, and 5. These wavelengths are shown in different bands in NIR reflectance spectra of wheat classes in Fig. 4.19. The top three principal components (2, 3, and 5) were selected. Based on the factor loadings, the top ten wavelengths of these principal components were identified.

Table 4.20 Top ten wavelengths of principal components 2, 3, and 5.

Wavelength No.	Principal component		
	2	3	5
1	1500	1700	1700
2	1510	1690	1690
3	1520	1680	1680
4	1490	1670	1320
5	1530	1070	1400
6	1450	1020	1670
7	1480	1050	1410
8	1580	960	1390
9	1470	1080	1350
10	1440	970	1140

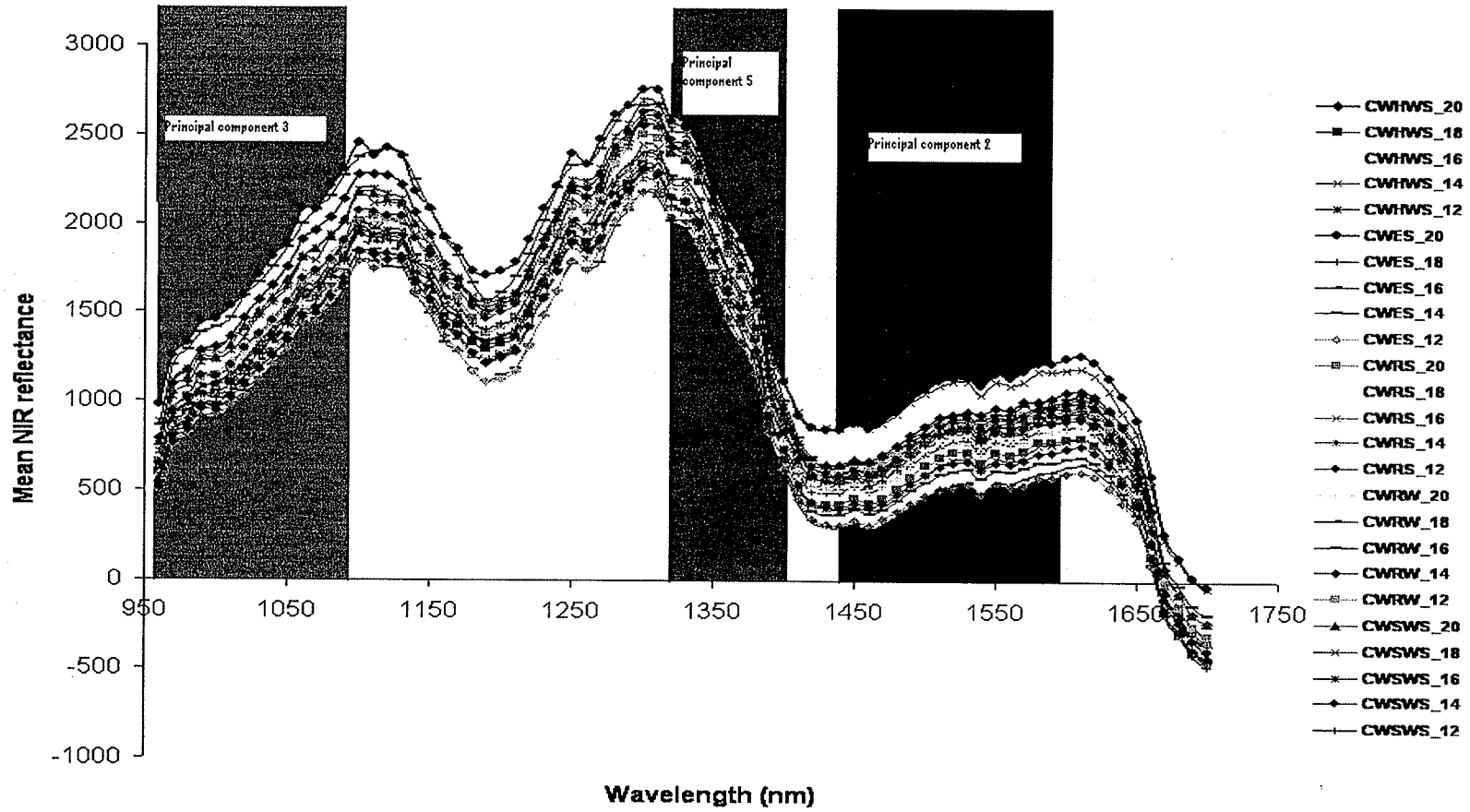

Fig. 4.19 Top wavelength bands of principal components 2, 3, and 5 in wheat class identification.

4.3 Prediction of protein and oil content in wheat (Study: 3)

4.3.1. Protein prediction in wheat

Measured protein contents (% dry basis) of the wheat classes are shown in Table 4.21. The protein content was the highest for CWRS wheat class followed by CWHWS, CWES, CWRW, and CWSWS wheat classes. The protein levels were higher than expected (usually 13 – 15%) because of a hot, dry harvest across western Canada in 2006.

Table 4.21 Measured protein contents (% dry basis) of wheat classes.

Replications	CWHWS	CWES	CWRS	CWRW	CWSWS
1	17.70	16.60	19.60	14.40	13.90
2	18.00	16.00	19.50	14.00	13.90
3	17.80	16.00	19.00	14.30	14.30
4	17.50	16.70	19.00	14.10	14.00
Mean	17.75	16.32	19.27	14.20	14.02
Standard deviation	0.21	0.38	0.32	0.18	0.19

4.3.1.1 PLSR model for protein prediction in wheat (75 input features)

The extracted PLS factors, their root mean PRESS, and their probability to exceed PRESS are given in Table 4.22. Minimum root mean PRESS value 0.30 was attained for the 12^{th} PLS factor. Also, for the 12^{th} PLS factor, the probability value to exceed PRESS was > 0.1. A twelve factor PLSR model was selected for predicting protein in wheat.

Table 4.22 Results of test set validation for the number of extracted factors of PLSR model for protein prediction in wheat (75 input features).

No. of extracted factors	Root mean PRESS	Probability > PRESS
1	0.77	< 0.0001
2	0.78	< 0.0001
3	0.46	< 0.0001
4	0.43	< 0.0001
5	0.40	< 0.0001
6	0.41	< 0.0001
7	0.39	< 0.0001
8	0.35	< 0.0001
9	0.33	< 0.0001
10	0.31	0.0350
11	0.31	0.001
12	0.30	1.0000
13	0.31	0.046
14	0.31	0.029
15	0.31	0.176

Table 4.23 shows the percent variation accounted for independent variables and dependant variable using the extracted PLS factors. This PLSR model could explain 98.29% of variations in the independent variables (NIR absorbance values) and 89.56% of variations in the dependant variable (protein).

Table 4.23 Percent variation accounted for by independent variables and dependant variable using PLS factors for protein prediction in wheat (75 input features)

No. of Extracted Factors	Model effects		Dependent variables	
	Current	Total	Current	Total
1	26.29	26.29	28.61	28.61
2	54.57	80.86	12.71	41.33
3	7.37	88.24	30.10	71.43
4	2.69	90.94	7.49	78.93
5	1.67	92.61	2.27	81.20
6	1.89	94.50	1.43	82.64
7	1.25	95.75	1.37	84.01
8	0.47	96.23	2.61	86.62
9	0.93	97.17	0.67	87.29
10	0.37	97.54	1.19	88.49
11	0.45	97.99	0.59	89.09
12	0.29	98.29	0.46	89.56

Fig. 4.20 shows the predicted and measured mean protein contents with their standard errors for the wheat classes. Standard deviations of predicted and measured protein contents of wheat were close to each other. Fig. 4.21 shows the scatter plot of predicted protein and measured protein contents of wheat samples. Performance of PLSR model could be explained from the value of RPD.

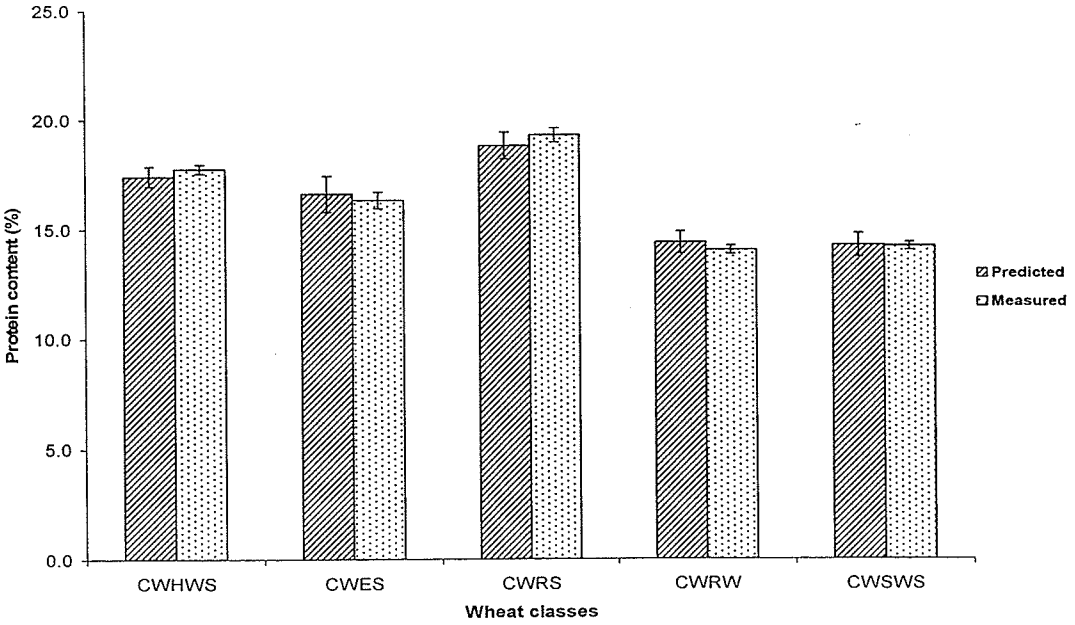

Fig. 4.20 Predicted and measured protein contents with their standard errors for wheat classes (75 input features).

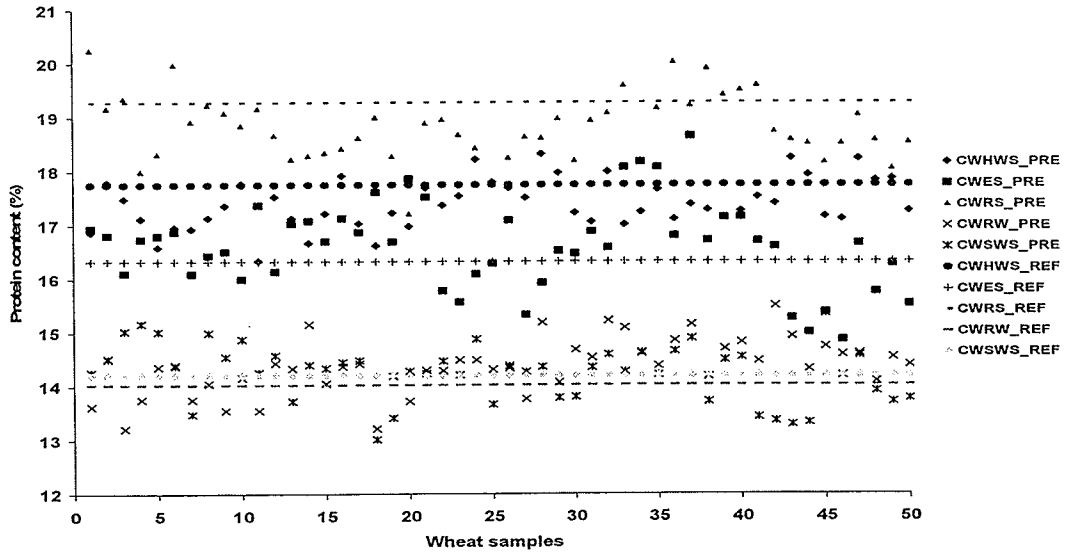

Fig. 4.21 Scatter plot of predicted protein content around measured mean values of protein for wheat classes (75 input features).

Predicted protein contents for wheat classes are listed in Table A.11. Results of grouping are shown in Table 4.24. Predicted protein contents of wheat classes were

grouped using PROC GLM in SAS. Scheffe's grouping, the most conservative grouping method, was used for grouping the predicted protein contents of wheat classes.

Table 4.24 Results of grouping for the predicted protein contents of wheat classes (75 input features).

Wheat class	Qualitative protein level	Predicted protein mean	Scheffe's test
CWRS	High	18.8	A
CWHWS	High	17.4	B
CWES	Medium	16.6	C
CWRW	Low	14.3	D
CWSWS	Low	14.2	D

The mean protein levels of CWRS, CWHWS, and CWES wheat classes were separated far from each other. The results of grouping showed that the predicted protein contents of high protein wheat classes (CWRS and CWHWS) and medium protein wheat class (CWES) were correctly grouped. Predicted protein contents of CWRS, CWHWS, and CWES wheat classes were significantly different from each other ($\alpha = 0.05$). But, predicted protein contents of low protein wheat classes such as CWRW and CWSWS were grouped in one group and for these wheat classes there was not significant difference ($\alpha = 0.05$) in protein content.

4.3.1.1.1 Statistical performance of PLSR model for protein prediction in wheat (75 input features)

Important statistical parameters (SEP, RPD, and RER) for the PLSR model are listed in Table 4.25. Standard deviation of the difference between the predicted and the measured values of any chemical compound is called standard error of performance (SEP). The ratio of standard error of performance to standard deviation of measured

values using a standard method is called RPD. Ratio of standard error of performance to the range of data in the standard method is called RER.

Table 4.25 Statistical performance of PLSR model for protein prediction in wheat (75 input features).

Parameters	Standard method	PLSR model	Difference
Overall Mean	16.31	16.28	0.02
SD	2.03	1.86	0.68 (SEP)
r		0.94	
RPD		2.98	
RER		7.71	

In the PLSR model with 75 input features, predicted protein had good correlation (r = 0.94) with measured protein. It showed that NIR hyperspectral imaging could be used for determining the protein content in wheat. RPD and RER values of PLSR model for protein were 2.98 and 7.71, respectively.

4.3.1.2 PLSR model for protein prediction in wheat (51 input features)

The extracted PLS factors, their root mean PRESS and their probability to exceed PRESS are given in Table 4.26. A minimum root mean PRESS value of 0.34 was obtained by the 9^{th} PLS factor. Also, for the 9^{th} PLS factor, the probability value to exceed PRESS was greater than 0.1. A nine factor PLSR model was used for predicting protein in wheat.

Table 4.26 Results of test set validation for the number of extracted factors of PLSR model for protein prediction in wheat (51 input features).

No. of extracted factors	Root mean PRESS	Probability > PRESS
1	0.79	< 0.0001
2	0.78	< 0.0001
3	0.46	< 0.0001
4	0.45	< 0.0001
5	0.42	< 0.0001
6	0.41	< 0.0001
7	0.38	< 0.0001
8	0.37	< 0.0001
9	0.34	1.0000
10	0.35	0.397
11	0.35	0.0810
12	0.35	0.0520
13	0.35	0.0960
14	0.36	0.0040
15	0.35	0.0310

Table 4.27 illustrates the percent variation accounted for by independent variables and dependant variable using the extracted PLS factors. PLSR model could explain 98.44% of variation in independent variables (NIR absorbance values) and 86.17% in the dependant variable (protein). The variation of independent variable explained by the 51 feature PLSR model was slightly more than that of the 75 feature PLSR model. But, percent variation in explaining the dependent variable by this model was around 3% less than that of the 75 feature PLSR model. The number of extracted factors was reduced from 12 to 9 and at the same time the percent variation in explaining the variation of protein content in wheat was also reduced by 3% than that of the 75 feature PLSR model.

Table 4.27 Percent variation accounted for by independent variables and dependant variable using PLS factors for protein prediction in wheat (51 input features).

No. of Extracted Factors	Model effects		Dependent variable	
	Current	Total	Current	Total
1	41.76	41.76	22.85	22.85
2	41.48	83.25	19.05	41.91
3	5.73	88.99	31.44	73.35
4	3.54	92.53	5.59	78.95
5	2.65	95.19	1.57	80.52
6	1.22	96.41	2.90	83.42
7	0.92	97.33	0.95	84.37
8	0.95	98.29	0.60	84.97
9	0.14	98.44	1.19	86.16

Fig. 4.22 explains predicted and measured protein contents and their standard errors of five wheat classes. Standard errors of predicted and measured protein contents were close to each other. Fig. 4.23 shows the scatter plot of predicted protein contents around the measured mean protein contents of wheat samples. This graph helps to visualize the variations of predicted and measured protein contents in wheat.

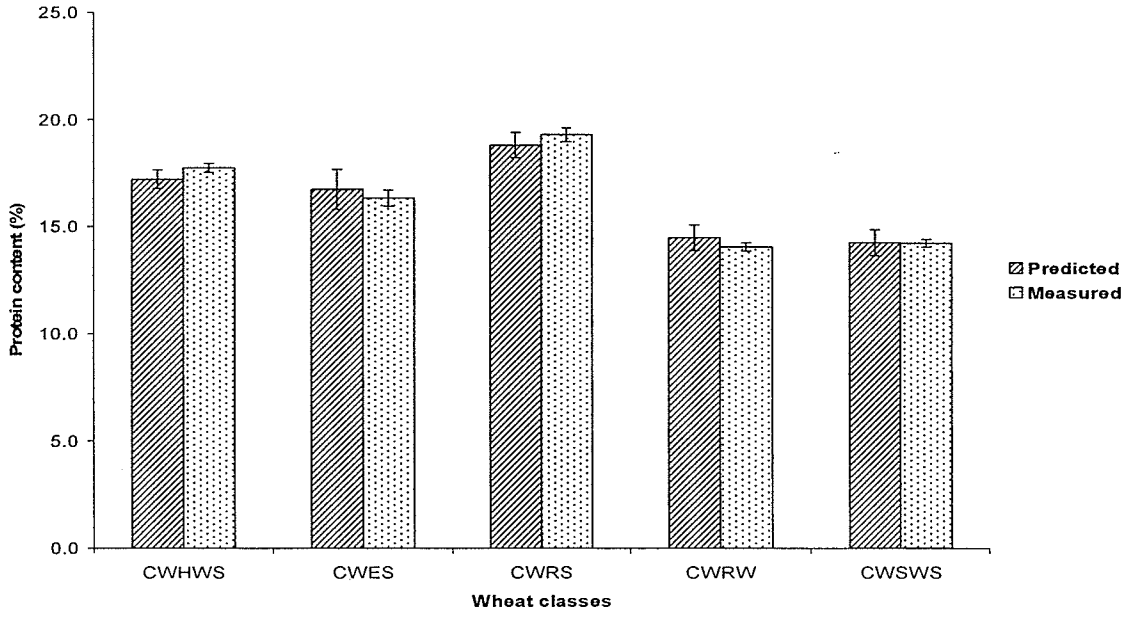

Fig. 4.22 Predicted and measured of protein content with their standard errors for wheat classes (51 input features).

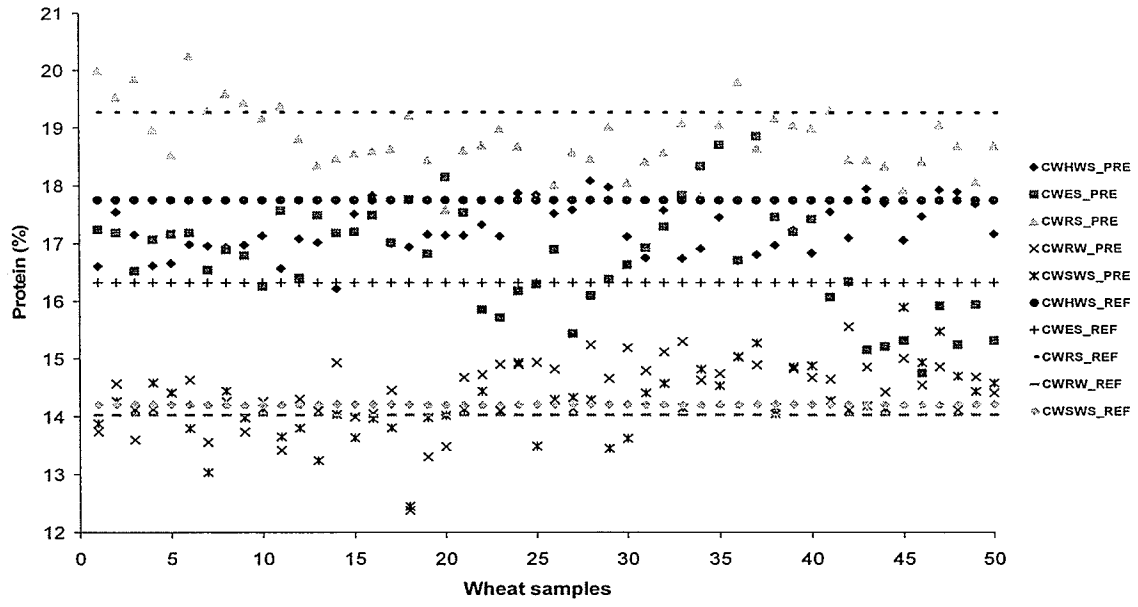

Fig. 4.23 Scatter plot of predicted protein content around the measured mean values of protein of wheat classes (51 input features).

Predicted protein values of wheat classes are listed in Table A.12. Results of grouping are shown in Table 4.28. Predicted protein contents of wheat classes were grouped using PROC GLM in SAS. Scheffe's grouping was used to group the predicted protein contents of wheat classes.

Table 4.28 Results of grouping for the predicted protein contents of wheat classes using PLSR model (51 input features).

Wheat class	Qualitative protein level	Predicted protein mean	Scheffe's test
CWRS	High	18.7	A
CWHWS	High	17.2	B
CWES	Medium	16.7	C
CWRW	Low	14.4	D
CWSWS	Low	14.2	D

The mean predicted protein contents of CWRS, CWHWS, and CWES wheat classes were separated far from each other. The results of grouping showed that the predicted protein contents of high protein wheat classes (CWRS and CWHWS) and a medium protein wheat class (CWES) were correctly grouped. The predicted protein contents of CWRS, CWHWS, and CWES wheat classes were significantly different from each other ($\alpha = 0.05$). But, predicted protein contents of low protein wheat classes (CWRW and CWSWS) were grouped into one group. The same trend was observed in the 75 feature PLSR model also. The predicted protein contents of CWRW and CWSWS wheat classes were not significantly different ($\alpha = 0.05$).

4.3.1.2.1 Statistical performance of PLSR model for protein prediction in wheat (51 input features)

Important statistical parameters (SEP, RPD, and RER) for the 51 feature PLSR model are listed in Table 4.29.

Table 4.29 Statistical performance of PLSR model for protein prediction in wheat (51 input features).

Parameters	Standard method	PLSR model	Difference
Overall Mean	16.31	16.28	0.02
SD	2.03	1.85	0.76 (SEP)
r		0.92	
RPD		2.64	
RER		6.82	

PLSR model for protein had a good correlation (r = 0.92) with the measured protein values. But, the correlation was less than that of the 75 feature PLSR model. It showed that NIR hyperspectral imaging could predict the protein content in wheat. A RPD value of 2.64 was obtained for PLSR model for protein. It was less than that of the 75 feature PLSR model. A RER value of 6.82 was observed for the 51 feature PLSR model for protein.

4.3.2 Oil content prediction in wheat

Measured oil contents (% dry basis) of the wheat classes are shown in Table 4.30. The oil content was the highest for CWRW wheat class followed by CWHWS, CWSWS, CWES, and CWRS wheat classes.

Table 4.30 Measured oil contents (% dry basis) of wheat classes.

Replications	CWHWS	CWES	CWRS	CWRW	CWSWS
1	1.70	1.36	1.36	1.88	1.72
2	1.56	1.40	1.30	1.82	1.70
3	1.70	1.38	1.42	1.80	1.62
Mean	1.65	1.38	1.36	1.83	1.68
Standard deviation	0.08	0.02	0.06	0.04	0.05

4.3.2.1 PLSR model for oil content prediction in wheat (75 input features)

The extracted PLS factors, their root mean PRESS values, and their probability to exceed PRESS in predicting oil content in wheat are given in Table 4.31. A minimum root mean PRESS value of 0.61 was obtained for the 13th PLS factor. For the 13th PLS factor, the probability value to exceed PRESS was greater than 0.1. A thirteen factor PLSR model was used for predicting oil content in wheat.

Table 4. 31 Results of test set validation for the number of extracted factors of PLSR model for oil content prediction in wheat (75 input features).

No. of extracted factors	(Root mean PRESS)	Probability > PRESS
1	1.01	< 0.0001
2	0.88	< 0.0001
3	0.72	< 0.0001
4	0.67	< 0.0001
5	0.68	< 0.0001
6	0.71	< 0.0001
7	0.67	< 0.0001
8	0.64	0.0060
9	0.67	< 0.0001
10	0.65	< 0.0001
11	0.66	< 0.0001
12	0.64	< 0.0001
13	0.61	1.0000
14	0.62	0.1540
15	0.63	< 0.0001

Table 4.32 illustrates the percent variation accounted for independent variables and a dependant variable using PLS factors to predict oil content in wheat. PLSR model

could explain 98.39% of variations in the independent variables (NIR absorbance values) and 67.99% of variations in the dependant variable (oil content).

Table 4.32 Percent variation accounted for by independent variables and dependant variable using PLS factors for oil content prediction in wheat (75 input features).

No. of Extracted Factors	Model effects		Dependent variable	
	Current	Total	Current	Total
1	58.59	58.59	8.82	8.82
2	22.32	80.91	19.87	28.70
3	7.30	88.22	18.11	46.81
4	2.65	90.88	5.38	52.19
5	1.93	92.82	1.56	53.76
6	1.73	94.55	1.56	55.33
7	1.06	95.62	1.77	57.10
8	0.82	96.45	2.00	59.11
9	0.73	97.18	2.26	61.37
10	0.28	97.46	3.03	64.40
11	0.41	97.88	0.81	65.22
12	0.39	98.27	0.82	66.05
13	0.11	98.39	1.94	67.99

When comparing this model with PLSR models for protein, the percent variation in explaining the dependant variable (oil content) was too low and approximately equal to 68% only. Fig. 4.24 explains predicted and measured oil contents and their variations in wheat. Standard errors of predicted and measured oil contents were close to each other for most of the wheat classes except for CWES wheat. This means that this PLSR model for oil content would poorly predict the oil content of CWES wheat. Fig. 4.25 shows the scatter plot of the predicted and measured oil contents of wheat samples.

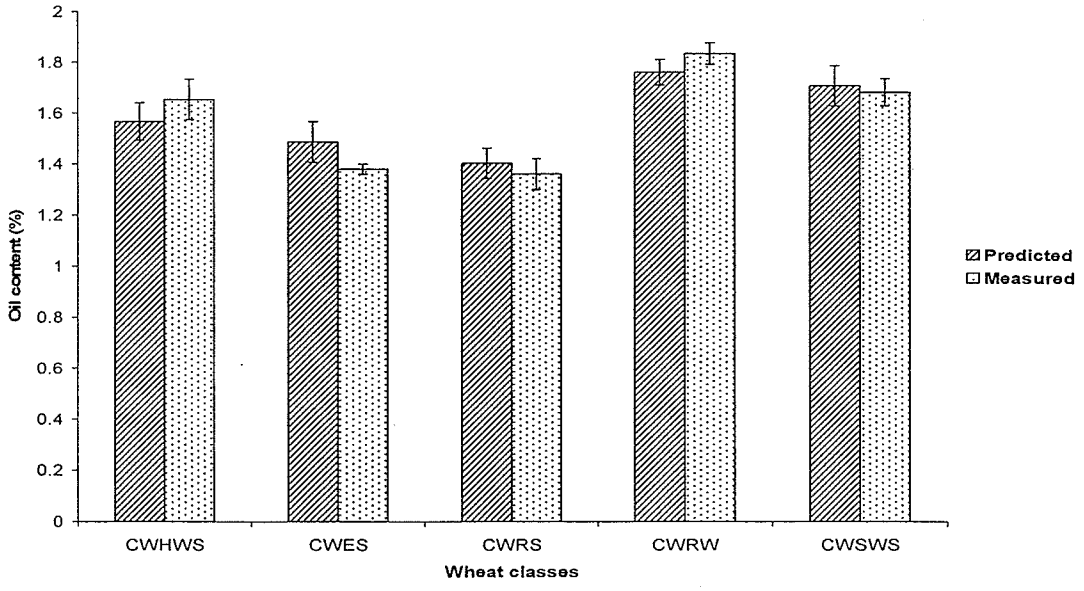

Fig. 4.24 Predicted and measured oil contents with their standard errors for wheat classes (75 input features).

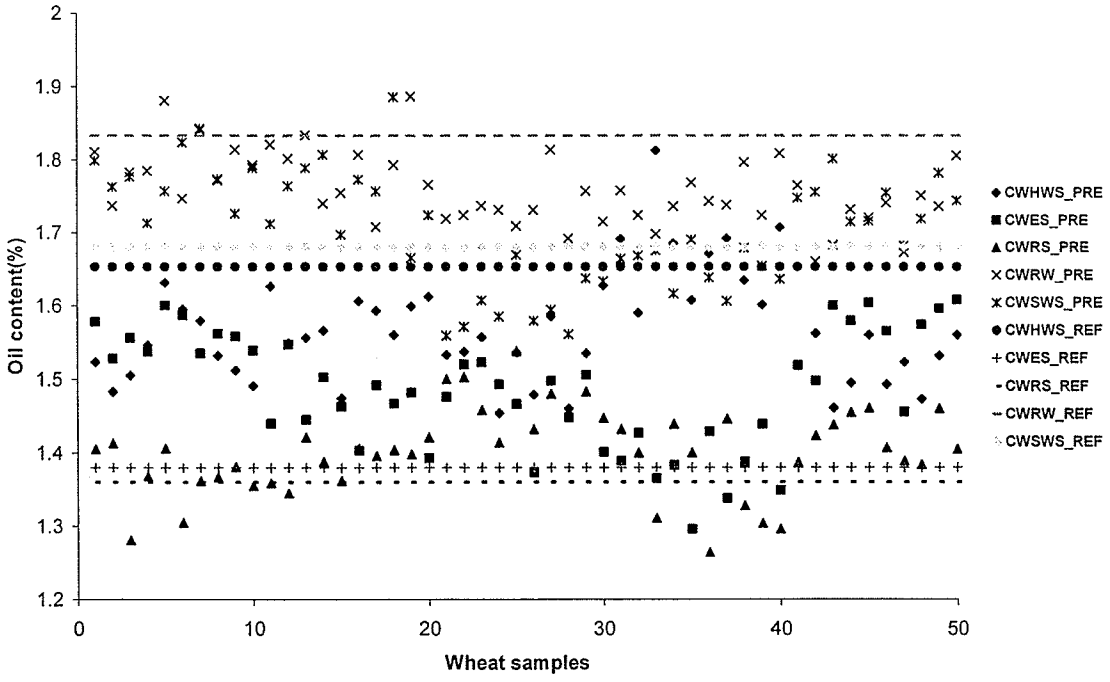

Fig. 4.25 Scatter plot of predicted oil content around measured mean values of oil content of wheat classes (75 input features).

Predicted oil contents of wheat classes are listed in Table A.13. Results of grouping are shown in Table 4.33. Predicted oil contents for five wheat classes were grouped using PROC GLM in SAS. Scheffe's grouping was used to group the predicted oil contents of wheat classes.

Table 4.33 Results of grouping for predicted oil contents of wheat classes using PLSR model (75 input features).

Wheat class	Qualitative protein level	Predicted oil content mean	Scheffe's test
CWRW	High	1.75	A
CWSWS	High	1.70	B
CWHWS	Medium	1.56	C
CWES	Low	1.48	D
CWRS	Low	1.40	E

The predicted oil contents of wheat classes were separated from each other and correctly grouped. It is shown that the predicted oil contents of the five wheat classes (CWRS, CWHWS, CWSWS, CWRW, and CWES) were significantly different from each other ($\alpha = 0.05$).

4.3.2.1.1 Statistical performance of PLSR model for oil content prediction in wheat (75 input features)

Important statistical parameters (SEP, RPD, and RER) of the 75 feature PLSR model for oil content are given in Table 4.34.

Table 4.34 Statistical performance of PLSR model for oil content prediction in wheat (75 input features).

Parameters	Standard method	PLSR model	Difference
Overall Mean	1.58	1.58	-0.002
SD	0.18	0.15	0.10 (SEP)
R		0.83	
RPD		1.82	
RER		4.70	

Predicted oil contents had good correlation (r = 0.83) with the measured oil contents of wheat classes. But, the value of r was less than that of the 75 and 51 feature PLSR models for protein in wheat. NIR hyperspectral imaging could be used for determining the oil content in wheat. A RPD value of 1.82 was obtained for PLSR model for oil content prediction. It was also less than that of the 75 and 51 feature PLSR models for protein prediction in wheat. A RER value of 4.70 was obtained for PLSR model for oil content.

4.3.2.2 PLSR model for oil content prediction in wheat (51 input features)

The extracted PLS factors, their root mean PRESS, and their probability to exceed PRESS are given in Table 4.35. A minimum root mean PRESS value of 0.63 was attained for the 11^{th} PLS factor. For the 10^{th} PLS factor, the probability value to exceed PRESS was greater than 0.1. A ten factor PLSR model was used to predict oil content in wheat.

Table 4.35 Results of test set validation for the extracted factors of PLSR model for oil content prediction in wheat (51 input features).

No. of extracted factors	Root mean PRESS	Probability > PRESS
1	1.00	< 0.0001
2	0.90	< 0.0001
3	0.74	< 0.0001
4	0.72	< 0.0001
5	0.72	< 0.0001
6	0.76	< 0.0001
7	0.69	< 0.0001
8	0.66	0.0010
9	0.66	< 0.0001
10	0.64	0.3430
11	0.63	1.0000
12	0.64	0.0410
13	0.64	0.1510
14	0.64	< 0.0001
15	0.64	0.1530

Table 4.36 illustrates the percent variation accounted for independent variables and a dependent variable using the extracted PLS factors to predict oil content in wheat. This PLSR model could explain 98.58% of variations in the independent variables (NIR absorbance values) and 62.88% in the dependent variable (oil content). Percent variation in explaining the dependent variable (here it is oil content) was approximately 5% less than that of the 75 feature PLSR model for oil content. At the same time, all PLSR models explained about 99% of variations in independent variables (NIR absorbance values).

Table 4.36 Percent variation accounted for by independent variables and dependant variable using PLS factors for oil content prediction in wheat (51 input features).

No. of Extracted Factors	Model effects		Dependent variable	
	Current	Total	Current	Total
1	46.19	46.19	12.95	12.95
2	36.62	82.81	13.96	26.91
3	6.15	88.97	19.49	46.41
4	3.51	92.49	3.76	50.18
5	2.75	95.24	0.98	51.16
6	1.40	96.65	2.13	53.30
7	1.22	97.87	1.75	55.05
8	0.38	98.26	3.27	58.33
9	0.14	98.40	2.53	60.86
10	0.17	98.57	2.12	62.87

When comparing with PLSR models for protein, the percent variation in explaining the dependent variable was far low and approximately equal to 63% only.

Fig. 4.26 explains predicted and measured oil contents and their standard errors in wheat. Standard deviations of the predicted and measured values of oil content were close to each other for most of the wheat classes except for CWES and CWRW wheat classes. It represented that PLSR model could poorly predict oil contents of CWRW and CWES wheat classes. Fig. 4.27 shows the scatter plot of predicted and measured oil contents of wheat samples.

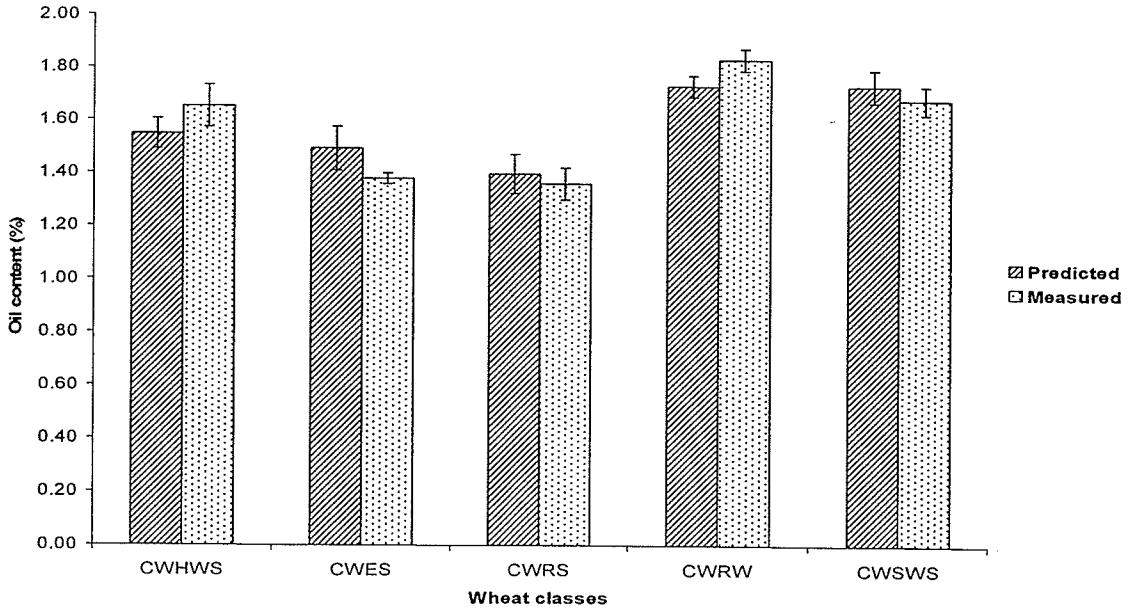

Fig. 4.26 Predicted and measured mean values of oil content with their standard errors for wheat classes (51 input features).

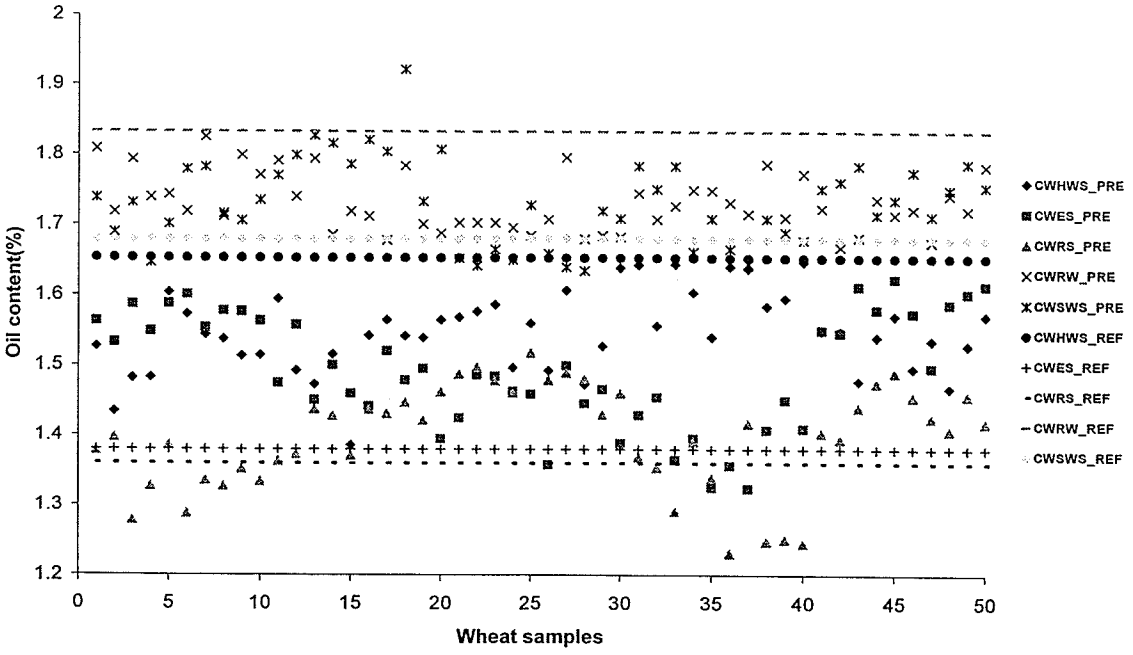

Fig. 4.27 Scatter plot of predicted oil contents around measured mean values of oil content for wheat classes (51 input features).

Predicted oil contents of wheat classes are listed in Table A.14. Results of grouping are shown in Table 4.37. Predicted oil contents of wheat classes were grouped using PROC GLM in SAS. Scheffe's grouping was used to group the predicted oil contents of wheat classes.

Table 4.37 Results of grouping for predicted oil content of wheat classes using PLSR model (51 input features).

Wheat class	Qualitative protein level	Predicted oil content mean	Scheffe's test
CWRW	High	1.73	A
CWSWS	High	1.73	A
CWHWS	Medium	1.54	B
CWES	Low	1.49	C
CWRS	Low	1.39	D

Predicted oil contents were grouped separately for the three wheat classes (CWHWS, CWES, and CWRS). Predicted oil contents were grouped in the same group for CWRW and CWSWS wheat classes. The predicted oil contents of CWRW and CWSWS wheat classes were close to each other and at the same time the predicted oil contents of CWHWS, CWES, and CWRS wheat classes were separated far from each other. Predicted oil contents of the three wheat classes (CWRS, CWHWS, and CWES) were significantly different from each other (α = 0.05) whereas they were not significantly different from each other (α = 0.05) for CWRW and CWSWS wheat classes.

4.3.2.2.1 Statistical performance of PLSR model for oil content prediction in wheat (51 input features)

Important statistical parameters (SEP, RPD, and RER) of the 51 feature PLSR model were found and listed in Table 4.38.

Table 4.38 Statistical performance of PLSR model for oil content prediction in wheat (51 input features).

Parameters	Standard method	PLSR model	Difference
Overall Mean	1.58	1.58	0.001
SD	0.18	0.14	0.10 (SEP)
r		0.80	
RPD		1.68	
RER		4.33	

Predicted oil contents of wheat classes had good correlation (r = 0.80) with the measured oil contents using the 51 feature PLSR model. But, the value of r was less than that of the 75 and 51 feature PLSR models for protein in wheat; and the 75 feature PLSR model for oil content in wheat. NIR hyperspectral imaging could be used to predict the oil content in wheat. A RPD value of 1.68 was obtained for the PLSR model. It was less than that of the 75 and 51 feature PLSR models for protein; and the 75 feature PLSR model for oil content in wheat. The RER value of PLSR model was 4.33.

Cogdill et al. (2004) found a SECV value of 1.20% and a RPD value of 2.74 for predicting the moisture content in maize. A SECV value of 1.38% and a RPD value of 1.45 were also obtained for oil content prediction in maize from the models developed using PLSR and PCR. Nagata et al. (2005) obtained a SEP value of 0.35 and a correlation of 0.78 for predicting the firmness of strawberries using three wavelength stepwise MLR model. A SEP value of 0.53 and a correlation of 0.87 were obtained for predicting SSC of strawberries using a five wavelength stepwise MLR model. Delwiche (1998) found that the NIR wavelength region of 1100 – 1400 nm was effective for protein content prediction in wheat. In error analysis, chemometric error was equal to 0.41% protein. Although, stepwise MLR models produced test set SEP values of 0.1 – 0.2% protein more than that of PLSR models, stepwise MLR was recommended for protein prediction

due to its less complexity. Wesley et al. (2001) found a SEP value of 0.65 using a curve fitting method and a SECV value of 0.38 using a PLSR model for glutenin content prediction in wheat flour. R^2 values of 0.85 – 0.93 and SEP values of 0.4 – 0.9% were reported for predicting protein in wheat using PLSR (Delwiche 1995).

The results of PLSR models for predicting protein and oil content in wheat showed that the statistical performances of the 75 feature PLSR models were better than that the 51 feature PLSR models. It was proved that the 75 feature PLSR models showed improved results in predicting protein and oil contents in wheat than that of the 51 feature PLSR models. NIR wavelengths of 960 – 1700 nm produced good results in predicting protein and oil contents in wheat and further reduction in upper and lower limits of the wavelength region would affect the performance of prediction. Environmental variables such as soil type, climate may affect the amount of protein and oil contents present in wheat in unpredictable ways and produce errors during prediction.

5. CONCLUSIONS

This study has established that NIR reflectance intensities and absorbance values of hyperspectral imaging could differentiate wheat classes and predict protein and oil contents of wheat. First, this study was used to identify the Canadian wheat classes at 11% moisture level and produced good results. The normalized NIR mean reflectance intensities of hyperspectral images could be used for differentiating wheat classes. Classification accuracies of 96 – 100% were obtained using LDA and QDA in wheat class identification. It was found that statistical classifier could be an effective statistical tool to differentiate wheat classes. In ANN, classification accuracies of wheat classes were usually > 90%, and only a few classification accuracies were 80 – 90%. ANN could also be effective when few distinctive wheat classes need to be classified. Classification accuracies of validation sets of a few wheat classes (CPSR, CWRW, and CWSWS) were 100%. Further research is needed to consider samples of different levels of moisture contents, different crop years, different locations and mixtures of two or more wheat classes at different levels to improve model robustness and classification accuracy.

Secondly, this study was used to differentiate wheat classes at different moisture levels. Mostly, LDA and QDA with 75, 51, and top 7 NIR absorbance features produced classification accuracies of > 90% for the wheat classes at various moisture levels. Canonical variable plots confirmed that the NIR absorbance values of hyperspectral images could be used to differentiate wheat classes based on the presence of moisture and protein contents in wheat. PCA and STEPDISC procedure were used to identify wavelengths responsible for wheat class identification. This study should be extended by considering the wheat samples of different crop years, different locations, and mixtures of

two or more wheat classes at different levels to improve model robustness and classification accuracy.

Finally, this study was used to predict protein and oil contents in wheat. NIR absorbance features at the wavelength region of 960 – 1700 nm were used to develop PLSR models and produced good results in predicting protein and oil contents in wheat. PLSR models for protein explained the variations of the dependent variable better than that of the PLSR models for oil content. NIR hyperspectral imaging could be used to measure protein and oil contents of wheat. Wheat samples from different crop years and different locations should be included to improve the model robustness and prediction accuracy.

6. REFERENCES

AACC Standards. 2000. Standard 30-25: Crude fat in wheat, corn, and soy flour, feeds, and mixed feeds. St. Paul, MN: AACC.

Anonymous. 1999a. The CANDISC procedure. In SAS/STAT user's guide, Version 8. Cary, NC: SAS Institute Inc.

Anonymous. 1999b. The PLS procedure. In SAS/STAT user's guide, Version 8. Cary, NC: SAS Institute Inc.

Anonymous. 2007a. SDS-PAGE (Polyacryamide gel electrophoresis). http: //www.davidson.edu/academic/biology/courses/Molbio/SDSPAGE/SDSPAGE.ht--ml (2007/07/20).

Anonymous. 2007b. Near-IR absorption bands. http://www.asdi.com/nir-chart_grid_rev-3.pdf (2007/05/10).

Anonymous. 2007c. Optimizing wheat yield and protein on the Canadian prairies. http://www.gov.mb.ca/agriculture/research/ardi/projects/images/98-147a.jpg (2007/05/09).

Anonymous. 2007d. How do halogen bulbs give constant light output? http://www .ushio.com/support/faqs.htm#7 (2007/07/31).

Anonymous. 2007e. Quadratic classifier. http://en.wikipedia.org/wiki/Quadratic_ classifier (2007/08/26).

Anonymous. 2007f. Principal components analysis. http://en.wikipedia.org/wiki/ Karhunen-Lo%C3%A8ve_transform (2007/08/26).

AOAC Standards. 2003. Standard 990.03: Protein (crude) in animal feed, combustion method in official methods of analysis of AOAC International. Gaithersburg, MD: AOAC.

Armstrong, P.R. 2006. Rapid single-kernel NIR measurement of grain and oil-seed attributes. *Applied Engineering in Agriculture* 22(5): 767-772.

ASAE Standards. 2003. Standard S352.2: Moisture measurement – unground grain and seeds. St. Joseph, MI: ASABE.

Baker, J.E., F.E. Dowell and J.E. Throne. 1999. Detection of parasitized rice weevils in wheat kernels with near infrared spectroscopy. *Biological Control* 16: 88-90.

Burger, J. and P. Geladi. 2006. Hyperspectral NIR imaging for calibration and prediction: a comparison between image and spectrometer data for studying organic and biological samples. *Analyst* 131: 1152-1160.

CGC. 2006a. Classes of Canadian wheat. http://www.grainscanada.gc.ca/Quality/ Wheat/classes-e.htm (2006/11/30).

CGC. 2006b. Kernel visual distinguishability in western Canadian wheat classes. Available at: http://www.grainscanada.gc.ca/quality/ wheat/classes_kvd-e.htm. Accessed on 25 July 2007.

CGC. 2007a. Variety identification monitoring. http://www.grainscanada.gc.ca/grl /variety_id/vid_monitoring-e.htm (2007/07/20).

CGC. 2007b. Western standards committee recommendations from April 11, 2007. http://www.grainscanada.gc.ca/Regulatory/Standards/western/ recommendations-07-04-11-e.htm#wheat (2007/07/25).

CWB. 2007. The wheat quality control system in Canada. http://www.cwb.ca/public/en /library/research/popups/wheat_Quality_standards.jsp (2007/07/24).

Cocchi, M., C. Durante, G. Foca, A. Marchetti, L. Tassi and A. Ulrici. 2006. Durum wheat adulteration detection by NIR spectroscopy multivariate calibration. *Talanta* 68: 1505-1511.

Cogdill, R.P., C.R. Hurburgh, Jr. and G.R. Rippke. 2004. Single-kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of ASAE* 47(1): 311-320.

Crow, G.H. 2006. Using SAS in Agricultural and Food Sciences Research. Winnipeg, MB: University of Manitoba.

Delwiche, S.R. 1995. Single wheat kernel analysis by near infrared transmittance: protein content. *Cereal Chemistry* 72(1): 11-16.

Delwiche, S.R. 1998. Protein content of single kernels of wheat by near infrared reflectance spectroscopy. *Journal of Cereal Science* 27: 241-254.

Delwiche, S.R. 2003. Classification of scab and other mold damaged wheat kernels by near-infrared reflectance spectroscopy. *Transactions of ASAE* 46(3): 731-738.

Delwiche, S.R. and D.R. Massie. 1996. Classification of wheat by visible and near-infrared reflectance from single kernels. *Cereal Chemistry* 73(3): 399-405.

Delwiche, S.R. and R.A. Graybosch. 2002. Identification of waxy wheat by near-infrared reflectance spectroscopy. *Journal of Cereal Science* 35: 29-38.

Dexter, J.E. and B.A. Marchylo. 2007. Recent trends in durum wheat milling and pasta processing: Impact on durum wheat quality requirements. http:// www.grainscanada.gc.ca/Pubs/confpaper/Dexter/trends/qualreq2-e.htm (2007/07/ 25).

Dowell, F.E. 2000. Differentiating vitreous and nonvitreous durum wheat kernels by using near-infrared spectroscopy. *Cereal Chemistry* 77(2): 155-158.

Dowell, F.E., T.C. Pearson, E.B. Maghirang, F. Xie and D.T. Wicklow. 2002. Reflectance and transmittance spectroscopy applied to detecting fumonisin in single corn kernels infected with *Fusarium verticillioides*. *Cereal Chemistry* 79(2): 222-226.

FAO. 2007. Food and Agriculture Organizations of the United Nations. http://www.fao.org/DOCREP/006/Y5022E/y5022e03.htm (2007/05/07).

FAOSTAT. 2006. Food and Agriculture Organizations of the United Nations. http://faostat.fao.org (2006/11/30).

Fowler, 2007. Grading and classes. http://www.usask.ca/agriculture/plantsci/winter_ cereals/Winter_wheat/CHAPT24/cvchpt24.php (2007/07/25).

Guy, R.C.E., B.G. Osborne and P. Robert. 1996. The application of near infrared reflectance spectroscopy to measure the degree of processing in extrusion cooking processes. *Journal of Food Engineering* 27: 241-258.

Hareland, G.A. 1994. Evaluation of flour particle size distribution by laser diffraction, sieve analysis and near infrared reflectance spectroscopy. *Journal of Cereal Science* 21: 183-190.

Heubner, F.R., T.C. Nelsen and J.A. Bietz. 1995. Differences among gliadins from spring and winter wheat cultivars. *Cereal Chemistry,* 72(4): 341-343.

ICC. 2001. Standard 143: Wheat-identification of varieties by electrophoresis. http://www.icc.or.at/6/wmethods.php#143 (2007/04/08).

Liu, M., S. Hu, H. Lin and E. Guo. 2007. Hyperspectral laser-induced fluorescence imaging for nondestructive assessing soluble solids content of orange. http://www.paper.edu.cn/downloadpaper.php?serial_number=200702-148 (2007/ 07/26).

Lu, R. 2003. Detection of bruises on apples using near-infrared hyperspectral imaging. *Transactions of ASAE* 46(2): 523-530.

Lu, R. and Y.R. Chen. 1998. Hyperspectral imaging for safety inspection of food and agricultural products. *Proceedings of SPIE* 3544: 120-133.

Maghirang, E.B., F.E. Dowell, J.E. Baker and J.E. Throne. 2003. Automated detection of single wheat kernels containing live or dead insects using near-infrared reflectance spectroscopy. *Transactions of ASAE* 46(4): 1277-1282.

Mehl, P.M., Y.R. Chen, M.S. Kim and D.E. Chen. 2004. Development of hyperspectral imaging technique for the detection of apple surface defects and contaminations. *Journal of Food Engineering* 61: 67-81.

Miralbes, C. 2004. Quality control in the milling industry using near infrared transmittance spectroscopy. *Food Chemistry* 88: 621-628.

Mohan, L.A., C. Karunakaran, D.S. Jayas and N.D.G. White. 2005. Classification of bulk cereals using visible and NIR reflectance characteristics. *Canadian Biosystems Engineering* 47, 7.7-7.14.

Murray, I. and P.C. Williams. 1987. Chemical principles of near-infrared technology. In: *Near-infrared Technology in the Agricultural and Food Industries,* eds. P.C. Williams and K.H. Norris, 17-34. St. Paul, MN: American Association of Cereal Chemists Inc.

Nagata, M., J.G. Tallada, T. Kobayashi and H. Toyoda. 2005. NIR hyperspectral imaging for measurement of internal quality in strawberries. Paper No. 053131. St. Joseph, MI: ASABE.

Nicolai, B.M., E. Lotze, A. Piers, N. Scheerlinck and K.I. Theron. 2006. Non-destructive measurement of bitter pit in apple fruit using NIR hyperspectral imaging. *Post harvest Biology and Technology* 40: 1-6.

Osborne, B.G., T. Fearn and P.H. Hindle. 1986. *Practical NIR Spectroscopy: With Applications in Food and Beverage Analysis.* 2nd ed. Essex, UK: Longman Scientific & Technical.

Osborne, B. G. 2006. Near-infrared Spectroscopy in Food Analysis. http://bsel.ist.utl.pt/2004/News/NIR%20Food%20Review.pdf (2006/04/20).

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

Paliwal, J., N.S. Visen and D.S. Jayas. 2001. Evaluation of neural network architectures for cereal grain classification using morphological features. *Journal of Agricultural Engineering Research* 79(4), 361-370.

Paliwal, J., W. Wang, S.J. Symons and C. Karunakaran. 2004. Insect species and infestation level determination in stored wheat using near-infrared spectroscopy. *Canadian Biosystems Engineering* 46: 7.17-7.24.

Pearson, T.C., D.T. Wicklow, E.B. Maghirang, F. Xie and F.E. Dowell. 2001. Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. *Transactions of ASAE* 44(5): 1247-1254.

Perez-Mendoza, J., J.E. Throne, F.E. Dowell and J.E. Baker. 2003. Detection of insect fragments in wheat flour by near-infrared spectroscopy. *Journal of Stored Products Research* 39: 305-312.

Petterson, H. and L. Aberg. 2003. Near infrared spectroscopy for determination of mycotoxins in cereals. *Food Control* 14: 229-232.

Ruan, R., Y. Li, X. Lin and P. Chen. 2002. Non-destructive determination of deoxynivalenol levels in barley using near-infrared spectroscopy. *Applied Engineering in Agriculture* 18(5): 549-553.

Rybicki. E. and M. Purves. 2007. SDS Polyacrylamide gel electrophoresis. http://www.mcb.uct.ac.za/sdspage.html (2007/07/20).

SAS. 2002. *SAS/STAT User's Guide.* Cary, N.C.: Statistical Analysis System, Inc. http://v8doc.sas.com/sashtml/stat/index.htm (2007/04/08).

Steinman, 2007. Wheat, Gluten allergy, Gluten intolerance and Gluten enteropathy. http://www.scienceinafrica.co.za/2001/december/gluten.htm (2007/07/20).

Teknomo, K. 2007. Linear discriminant analysis. http://people.revoledu.com/kardi/tutorial/LDA/LDA.html#LDA (2007/08/26).

Wang, D., F.E. Dowell and R.E. Lacey. 1999. Single wheat kernel color classification by using near-infrared reflectance spectra. *Cereal Chemistry* 76(1): 30-33.

Wang, D., F.E. Dowell and D.S. Chung. 2001. Assessment of heat-damaged wheat kernels using near-infrared spectroscopy. Paper No. 01-6006. St. Joseph, MI: ASABE.

Wang, D., F.E. Dowell and R. Dempster. 2002. Determining vitreous subclasses of hard red spring wheat using visible/near-infrared spectroscopy. *Cereal Chemistry* 79(3): 418-422.

Wang, D., M. S. Ram and F.E. Dowell. 2002. Classification of damaged soybean seeds using near-infrared spectroscopy. *Transactions of ASAE* 45(6): 1943-1948.

Wang, D., F.E. Dowell, M.S. Ram and W.T. Schapaugh. 2004. Classification of fungal-damaged soybean seeds using near-infrared spectroscopy. *International Journal of Food Properties* 7: 75-82.

Wang. W., J. Paliwal and D.S. Jayas. 2004. Determination of moisture content of ground wheat using near-infrared spectroscopy. Paper no: MB04-200. St. Joseph, MI: ASABE.

Wang, W and J. Paliwal. 2005. Correction of curvature-induced spectral variability in hyperspectral images of wheat kernels. Paper No: 053070. St. Joseph, MI: ASABE.

Wesley, I.J., O. Larroque, B.G. Osborne, N. Azudin, H. Allen and J.H. Skerritt. 2001. Measurement of Gliadin and Glutenin content of flour by NIR spectroscopy. *Journal of Cereal Science* 31: 125-133.

Williams, P., 2006. Marketing wheat by electronics. http://www. pdkprojects.com/ pdf/Grain%20Kernels%20No%203.pdf (2007/07/25).

7. APPENDIX

Literature review tables and confusion matrices

7.1 Overview of classification and prediction methods used in NIR studies

LDA, QDA, ANN, Canonical discriminant analysis, PLSR, and PCA have already been explained in Materials and Methods section of the thesis. Some of the other important methods used in NIR studies are briefly discussed here.

Multiple linear regression (MLR)

In this method, regression model is developed using linear combinations of two or more independent variables to explain a dependent variable.

$$y = a + bx_1 + cx_2 + dx_3$$

Where, y is a dependent variable; a is an intercept; b, c, and d are regression coefficients; x_1, x_2, and x_3 are independent variables.

Wavelet interface to linear modeling analysis (WILMA)

It is based on combination of fast wavelet transform (FWT) with MLR and PLSR techniques to develop regression models using independent variables to explain dependent variables. In this technique, each independent variable is decomposed into FWT variables and then those variables are used to develop regression models using MLR or PLSR methods (Cocchi et al. 2006).

Soft independent modeling of class analogy (SIMCA)

It is an approach in which PLSR is used to develop local models for each class and further the classes of new observations are predicted using those models (Delwiche 2003).

Table A.1 Comparison of materials, wavelength ranges and intervals, and quality parameters of agricultural and food products used in spectroscopy.

Material	Wavelength range (interval)	Property	Reference
NIR reflectance spectroscopy			
Ground wheat	850 – 2000 nm (5 nm)	Moisture content	Wang et al. (2004)
Soybean	400 – 1700 nm (5 nm)	Different types of damages in soybean	Wang et al. (2002)
Soybean	400 – 1700 nm (5 nm)	Classification of fungal damage	Wang et al. (2004)
Wheat	1100 – 2498 nm (2 nm)	Identification of waxy wheat	Delwiche and Graybosch (2002)
Wheat	1002 – 1704 nm (6 nm)	Scab and mold damage	Delwiche (2003)
Wheat	400 – 1700 nm	Heat damage in wheat	Wang et al. (2001)
Wheat	1100 – 2498 nm (2 nm)	Protein content	Delwiche (1998)
Wheat	400 – 1700 nm (5 nm)	Detection of live or dead internal rice weevil	Maghirang et al. (2003)
Wheat flour	400 – 2500 nm	Particle size distribution	Hareland (1994)
Wheat flour	550 – 1700 nm	Detection of insect fragments	Perez-Mendoza et al. (2003)
Wheat flour	400 – 2498 nm (2 nm)	Gliadin and glutenin content	Wesley et al. (2001)
Wheat flour	400 – 2498 nm (2 nm)	Adulteration in wheat flour	Cocchi et al. (2006)
Wheat starch, Wheat flour, Wholemeal flour	1100 – 2500 nm (4 nm)	Starch structural changes	Guy et al. (1996)
NIR transmittance spectroscopy			
Wheat	570 – 1100 nm	Mycotoxins (DON levels)	Petterson and Aberg (2003)
Wheat	850 – 1050 nm	Protein content	Delwiche (1995)
Wheat flour	850 – 1050 nm (2 nm)	Protein, moisture, dry gluten, wet gluten, starch damage and ash content	Miralbes (2004)
NIR spectroscopy (absorbance)			
Barley	400 – 2500 nm (2 nm)	DON levels	Ruan et al. (2002)
Wheat	400 – 1700 nm (5 nm)	Vitreousness of wheat	Dowell (2000)

Wheat	400 – 1700 nm	Detection of parasitized rice weevils	Baker et al. (1999)
Vis-NIR reflectance spectroscopy			
Wheat	551 – 750 nm and 1100 – 2498 nm	Identification of wheat classes	Delwiche and Massie (1996)
Wheat	400 – 1700 nm	Dark hard vitreousness of wheat	Wang et al. (2002)
Vis-NIR transmittance and reflectance spectroscopy			
Corn	550 – 1050 nm (transmittance) 400 – 1700 nm (reflectance)	Detection of fumonisin levels	Dowell et al. (2002)
Corn	500 – 900 nm (transmittance) 550 – 1700 nm (reflectance)	Detection of aflatoxin	Pearson et al. (2001)
NIR hyperspectral imaging			
Apple	900 – 1700 nm	Detection of bruises	Lu (2003)
Apple	900 – 1700 nm	Bitter pit lesions	Nicolai et al. (2006)
Maize	750 – 1090 nm (5 nm) (transmittance mode)	Moisture and oil content	Cogdill et al. (2004)
Strawberry	650 – 1000 nm (5 nm)	Firmness and soluble solids content	Nagata et al. (2005)
Wheat	1100 – 1600 nm (30 evenly spaced intervals)	Correction of curvature induced spectral variability	Wang and Paliwal (2005)
Vis-NIR hyperspectral imaging			
Apple	430 – 900 nm	Surface defects	Mehl et al. (2004)

Table A.2 Comparison of models used and results found to predict the quality parameters of agricultural products used in spectroscopy.

Goals in different crops	Models used	Results (Data pretreatment, if any)	Reference
PLSR and other main models used in NIR reflectance spectroscopic studies			
Soybean			
Different types of damages	PLSR	> 97% classification accuracy for two class models (sound vs. damaged) Average classification accuracies of 65 – 76% for six class models (sound / weather / frost / sprout / heat / mold damaged) at three different wavelength regions (490 – 750 nm, 490 – 1690 nm, and 750 – 1690 nm)	Wang et al. (2002)
	ANN	Average classification accuracies of 50 – 89% for six class models using three different hidden layer nodes (0, 3, 6, and 25) Average classification accuracies of 85 – 92% for six class models using three different momentum (0.4, 0.5, and 0.6) and learning rates (0.5, 0.6, and 0.7) Average classification accuracies of 66 – 93% for six class models at three different wavelength regions	
Classification of fungal damage	PLSR	> 97% classification accuracy for two class models (sound vs. damaged)	Wang et al. (2004)
	ANN	Average classification accuracies of 83 – 95% for five-class ANN models (healthy / phomopsis / *C. kikuchii* / SMV / downy mildew)	

Wheat			
Protein content	PLSR	R^2 values of 0.901 – 0.979 for five wheat classes (HRW / HRS / SRW / HWW / SWW) during model calibration Standard error of calibration of 0.389 – 0.518, Standard error or validation of 0.418 – 0.589, Standard error of test of 0.462 – 0.590 for five wheat classes	Delwiche (1998)
	MLR	R^2 values of 0.900 – 0.960 for five wheat classes during eight term MLR model calibration Standard error of calibration of 0.513 – 0.623, Standard error of validation of 0.566 – 0.758, Standard error of test of 0.618 – 0.720 for five wheat classes	
Assessment of heat damage	PLSR	> 98% classification accuracy for two class models (undamaged vs. damaged)	Wang et al. (2001)
	MLR	Classification accuracies of 91.8 – 99.3% for two class two wavelength MLR models	
Detection of live or dead internal rice weevil			
Day 1 (live)	PLSR	R^2 = 0.22 – 0.74, SECV = 0.21 – 0.44, Correct classification (CC) = 71 – 95%	Maghirang et al. (2003)
Day 7 (dead)	PLSR	R^2 = 0.02 – 0.78, SECV = 0.23 – 0.50, CC = 68 – 95%	
Day 14 (dead)	PLSR	R^2 = 0.08 – 0.78, SECV = 0.23 – 0.49, CC = 62 – 96%	
Day 28 (dead)	PLSR	R^2 = 0.08 – 0.72, SECV = 0.27 – 0.48, CC = 62 – 94%	
Day 42 (dead)	PLSR	R^2 = 0.03 – 0.76, SECV = 0.24 – 0.50, CC = 57 – 95%	
Day 56 (dead)	PLSR	R^2 = 0.04 – 0.74, SECV = 0.26 – 0.49, CC = 56 – 95%	

Wheat flour			
Glutenin	PLSR	$R^2 = 0.83$, SECV = 0.38	Wesley et al. (2001)
Gliadin	PLSR	$R^2 = 0.78$, SECV = 0.43	
Glutenin	Curve fitting	$R^2 = 0.71$, SEP = 0.65	
Gliadin	Curve fitting	$R^2 = 0.46$, SEP = 1.02	
Flour particle size distribution			
10 μm	PLSR	SEC = 0.16, $R^2 = 0.99$, SECV = 0.26	Hareland (1994)
10 – 41 μm	PLSR	SEC = 0.59, $R^2 = 0.99$, SECV = 0.87	
41 – 300 μm	PLSR	SEC = 0.75, $R^2 = 0.99$, SECV = 1.11	
Adulteration	PLSR	RMSEC = 0.29, RMSECV = 0.72, RMSEP = 0.39 (SNV pretreated data)	Cocchi et al. (2006)
	WILMA-MLR	RMSEC = 0.57, RMSECV = 0.68, RMSEP = 0.56	
	WILMA-PLS	RMSEC = 0.32, RMSECV = 0.62, RMSEP = 0.44	
Detection of insect fragments			
75 insect fragments or more	PLSR	Classification accuracy = > 90%	Perez-Mendoza et al. (2003)
Less than 75 insect fragments	PLSR	Classification accuracy = 20 – 40%	
130 insect fragments or more	PLSR	Classification accuracy = 90%	
Less than 130 insect fragments	PLSR	Classification accuracy = 83.3%	

PCA and other main models used in NIR reflectance spectroscopic studies

Ground wheat			
Moisture content	Linear calibration models	$R^2 = 0.793$, RMSEC = 0.541 (Averaged spectra) $R^2 = 0.972$, RMSEC = 0.239 (First derivatives of averaged spectra)	Wang et al. (2004)

Wheat			
Identification of waxy wheat			
Year 1	PCA, Statistical classifier	LDA = Classification accuracy of 42.3 – 71.2% (using 1 – 10 top principal component scores) QDA = Classification accuracy of 51.6 – 63.3% (using 1 – 10 top principal component scores)	Delwiche and Graybosch (2002)
Year 2	PCA, Statistical classifier	LDA = Classification accuracy of 47.2 – 68.2% (using 1 – 10 top principal component scores) QDA = Classification accuracy of 46.5 – 71.7% (using 1 – 10 top principal component scores)	
Identification of scab and mold damage	Ratios of wavelengths	Classification accuracy of cross validation = 68.3 – 95.7% (precise orientation) Classification accuracy of test set = 68.3 – 95.7% (precise orientation) Classification accuracy of cross validation = 75 – 92.3% (random orientation) Classification accuracy of test set = 64.8 – 88.9% (random orientation)	Delwiche (2003)
	PCA	Classification accuracy of cross validation = 89.3% (precise orientation) Classification accuracy of test set = 86.4% (precise orientation) Classification accuracy of cross validation = 84.7% (random orientation) Classification accuracy of test set = 86.4% (random orientation)	
	SIMCA-PCA	Classification accuracy of cross validation = 85.3 – 86.7% (precise orientation) Classification accuracy of test set = 83.6 – 85.8% (precise orientation) Classification accuracy of cross validation = 77.3 – 83.3% (random orientation)	

		Classification accuracy of test set = 73.4% - 83.7% (random orientation)	
Wheat flour			
Specific mechanical energy in starch during extrusion cooking			
Wheat starch	PCA and	R = 0.94, SEC = 39 kJ/kg, SEP = 51 kJ/kg	Guy et al. (1996)
Wheat flour	Forward	R = 0.98, SEC = 45 kJ/kg, SEP = 45 kJ/kg	
Wholemeal	stepwise regression	R = 0.99, SEC = 11 kJ/kg, SEP = 16 kJ/kg	

PLSR models used in NIR transmittance spectroscopic studies

Wheat

Determination of Mycotoxins			
Nordic material			
850 – 1050 nm	PLSR	R = 0.955, SECV = 638 µg DON / kg	Petterson and Aberg (2003)
570 – 1100 nm	PLSR	R = 0.978, SECV = 459 µg DON / kg	
670 – 1100 nm	PLSR	R = 0.984, SECV = 381 µg DON / kg	
Austrian dilution series			
850 – 1050 nm	PLSR	R = 0.963, SECV = 761 µg DON / kg	
570 – 1100 nm	PLSR	R = 0.891, SECV = 1292 µg DON / kg	
670 – 1100 nm	PLSR	R = 0.899, SECV = 1234 µg DON / kg	
Protein content (%)	PLSR	R^2 = 0.782 – 0.914, SEP = 0.42 – 0.94 (no pretreatment) R^2 = 0.872 – 0.918, SEP = 0.48 – 0.83 (MSC pretreated data) R^2 = 0.889 – 0.938, SEP = 0.45 – 0.83 (MSC pretreated on 2nd derivatives)	Delwiche (1995)

Wheat flour

		SEC	SECV	SEP	R^2	
Protein (%)	PLSR	0.11	0.12	0.14	0.99	Miralbes (2004)
Moisture (%)	PLSR	0.12	0.13	0.15	0.99	

Wet gluten (%, 14% mb)	PLSR	0.66	0.66	0.86	0.96
Dry gluten (%, 14% mb)	PLSR	0.17	0.17	0.22	0.99
Damage starch (Corrected Chopin Dubois units)	PLSR	1.01	1.25	1.63	0.94
Ash (%)	PLSR	0.023	0.021	0.024	0.98
Stability (min)	PLSR	0.76	0.79	1.02	0.88
Degree of dough softening in 20 min mixing (BU)	PLSR	6.4	6.74	8.77	0.93
Degree of dough softening in 12 min after development time (BU)	PLSR	6.2	6.6	11.0	0.90
Farinograph quality number	PLSR	7.3	7.7	9.1	0.92
Resistance of dough to deformation (mm)	PLSR	4.44	4.67	6.07	0.90
Ratio of deformation	PLSR	0.07	0.07	0.04	0.79
Deformation energy (10^{-4} J)	PLSR	15.6	16.5	21.5	0.95
		(Data are pretreated using SNT + DET procedures, and transformed with first derivative processing)			

PLSR and other main models used in NIR spectroscopic (absorbance) studies

Barley

DON levels (ppm)		R^2	SEP	
400 – 2500 nm (2 nm)	ANN	0.933	3.097	Ruan et al. (2002)
400 – 2500 nm (4 nm)	ANN	0.923	3.431	
400 – 2500 nm (10 nm)	ANN	0.924	3.490	
400 – 2500 nm (20 nm)	ANN	0.915	3.694	
400 – 2500 nm (40 nm)	ANN	0.921	3.474	
400 – 700 nm (10 nm)	ANN	0.921	3.351	
700 – 1100 nm (10 nm)	ANN	0.912	3.706	
1100 – 2500 nm (10 nm)	ANN	0.840	4.958	
1500 – 1850 nm (10 nm)	ANN	0.805	5.461	

Wheat			
Vitreousness			
All kernels	PLSR	R^2 = 0.26, Correct classification rate = 75% (cross validation), 73% (calibration), 72% (prediction)	Dowell (2000)
Obvious vitreous or non vitreous kernels	PLSR	R^2 = 0.85, Correct classification rate = 99.8% (cross validation), 100% (calibration), 100% (prediction)	
Detection of parasitized rice weevils			Baker et al. (1999)
Infested vs. uninfested	PLSR	R^2 = 0.90, SECV = 0.15	
Parasitoid vs. weevil	PLSR	R^2 = 0.87, SECV = 0.18	

PLSR and other main models used in Vis/NIR reflectance spectroscopic studies

Wheat			
Identification of classes			
Two-class models			
Visible region	PLSR	Average classification accuracy = 72.72 – 99.01%	Delwiche and Massie (1996)
NIR region	PLSR	Average classification accuracy = 78.39 – 92.38%	
Red vs. white wheat classes	PLSR	Classification accuracy = 97.5% (white), 98.8% (red)	
Interwhite wheat classes	PLSR	Classification accuracy = 89.3% (HWH), 93.6% (SWH)	
Interred wheat classes	PLSR	Classification accuracy = 92.0% (hard red wheat	

		classes), 73.4% (HRS), 85.9% (HRW), 65.5% (SRW)	
Red vs. white wheat classes	MLR	Classification accuracy = 96.5% (white), 98.2% (red)	
Interwhite wheat classes	MLR	Classification accuracy = 86.8% (HWH), 92.1% (SWH)	
Interred wheat classes	MLR	Classification accuracy = 91.6% (hard red wheat classes), 75.8% (HRS), 78.1% (HRW), 63.3% (SRW)	
Determination of dark hard vitreousness			
Including bleached kernels	PLSR	Average classification accuracy = 90.7 – 97.1% (cross validation) Average classification accuracy = 90.7 – 97.1% (test)	Wang et al. (2002)
Excluding bleached kernels	PLSR	Average classification accuracy = 97.1 – 100.0% (cross validation) Average classification accuracy = 97.5 – 100.0% (test)	

PLSR and other main models used in Vis/NIR transmittance and reflectance spectroscopic studies

Corn

Detection of fumonisin levels			
1 – 10 ppm	PLSR,	0.6 – 7.2% error	Dowell et al. (2002)
10 – 100 ppm	Discriminant	23.5 – 73.0% error	
>100 ppm	analysis	0 – 1.7% error	
Detection of aflatoxin			
0 ppb	Discriminant	0 – 0.6% error	Pearson et al. (2001)
1 – 10 ppb	analysis	1.0 – 3.1% error	
10 – 100 ppb		52.2 – 87.0% error	
>100 ppb		2.4 – 4.8% error	

0 ppb	PLSR	0 – 2.6% error
1 – 10 ppb		2.1 – 4.7% error
10 – 100 ppb		78.9 – 91.7% error
>100 ppb		4.9 – 14.7% error

Main models in NIR hyperspectral imaging studies

Apple

Detection of bruises	Principal components transform, minimum noise transform	Bruise detection accuracy = 62 – 88% (Red delicious apples) Bruise detection accuracy = 59 – 94% (Golden delicious apples)	Lu (2003)
Detection of bitter pit lesions	Discriminant PLS model	Pixels occupied by bitter pit lesions were separated from the pixels of healthy skin	Nicolai et al. (2006)

Maize

Moisture		R^2	SECV	
None (No outliers removed)	PLSR	0.786	1.91	
None (outliers removed)	PLSR	0.871	1.20	Cogdill et al. (2004)
SNV pretreated	PLSR	0.836	1.36	
DET pretreated	PLSR	0.872	1.20	
SNV – DET pretreated	PLSR	0.828	1.35	
MSC pretreated	PLSR	0.664	1.95	
None (No outliers removed)	PCR	0.558	2.31	
None (outliers removed)	PCR	0.856	1.25	
SNV pretreated	PCR	0.786	1.52	
DET pretreated	PCR	0.858	1.25	
SNV – DET pretreated	PCR	0.792	1.50	
MSC pretreated	PCR	0.647	2.01	
None (No outliers removed)	Genetic	0.795	1.55	

None (outliers removed)	algorithm + MLR	0.853	1.27	
SNV pretreated	GA+MLR	0.745	1.69	
DET pretreated	GA+MLR	0.820	1.36	
SNV – DET pretreated	GA+MLR	0.806	1.46	
Oil content				
None (No outliers removed)	PLSR	0.385	1.88	
None (outliers removed)	PLSR	0.515	1.40	
SNV pretreated	PLSR	0.538	1.38	
DET pretreated	PLSR	0.491	1.47	
SNV – DET pretreated	PLSR	0.557	1.37	
None (No outliers removed)	PCR	0.114	2.05	
None (outliers removed)	PCR	0.470	1.45	
SNV pretreated	PCR	0.467	1.43	
DET pretreated	PCR	0.419	1.49	
SNV – DET pretreated	PCR	0.492	1.42	
None (No outliers removed)	GA+MLR	0.439	1.62	
None (outliers removed)	GA+MLR	0.508	1.41	
SNV pretreated	GA+MLR	0.454	1.50	
DET pretreated	GA+MLR	0.407	1.56	
SNV – DET pretreated	GA+MLR	0.554	1.39	
Strawberry				
Firmness				
70% - Full-ripe	MLR	SEP = 0.241 – 0.262, R = 0.588 – 0.645		Nagata et al. (2005)
50% - Full-ripe	MLR	SEP = 0.344 – 0.350, R = 0.786 – 0.796		
Soluble solids content				
70% - Full-ripe	MLR	SEP = 0.43 – 0.80, R = 0.32 – 0.87		
All ripeness levels	MLR	SEP = 0.58 – 0.69, R = 0.59 – 0.73		

Wheat			
Correction of curvature induced spectral variability	Morphological shrinking, Simulated ellipsoidal surface fitting	Morphological shrinking was more effective. Both methods performed well in reducing the spectral variability along the minor axis of the endosperm of the wheat kernel. Ellipsoidal surface fitting was not effective in correcting the spectral variability along the major axis of the endosperm of wheat kernel.	Wang and Paliwal (2005)
Main models in Vis/NIR hyperspectral imaging studies			
Apple			
Identification of surface defects	Asymmetric and symmetric second difference models	685 nm (Chlorophyll absorption wavelength) 722 nm and 865 nm (NIR wavelengths) were responsible for detection of surface defects in apples	Mehl et al. (2004)

Table A.3 Confusion matrix for the 75 NIR reflectance feature LDA model for eight wheat classes (n = 300 per class)

Class (To)→ / (From)↓	CPSR	CPSW	CWAD	CWES	CWHWS	CWRS	CWRW	CWSWS
CPSR	300	0	0	0	0	0	0	0
CPSW	0	296	1	0	0	3	0	0
CWAD	0	2	296	2	0	0	0	0
CWES	0	0	0	300	0	0	0	0
CWHWS	0	0	0	0	300	0	0	0
CWRS	0	0	0	0	0	300	0	0
CWRW	0	0	0	0	0	0	300	0
CWSWS	0	0	0	0	0	0	0	300

Table A.4 Confusion matrix for the 75 NIR reflectance feature QDA model for eight wheat classes (n = 300 per class)

Class (To) → / (From) ↓	CPSR	CPSW	CWAD	CWES	CWHWS	CWRS	CWRW	CWSWS
CPSR	299	1	0	0	0	0	0	0
CPSW	0	291	5	0	0	4	0	0
CWAD	0	1	298	1	0	0	0	0
CWES	0	8	1	290	1	0	0	0
CWHWS	0	0	0	0	296	4	0	0
CWRS	0	5	0	0	0	295	0	0
CWRW	0	0	1	0	0	0	299	0
CWSWS	2	1	1	0	0	0	4	292

7.2 List of wheat classes and their moisture levels for Tables A.5 to A.10

S. No.	Wheat class and its moisture content
1	CWES wheat at 12% moisture level
2	CWES wheat at 14% moisture level
3	CWES wheat at 16% moisture level
4	CWES wheat at 18% moisture level
5	CWES wheat at 20% moisture level
6	CWHWS wheat at 12% moisture level
7	CWHWS wheat at 14% moisture level
8	CWHWS wheat at 16% moisture level
9	CWHWS wheat at 18% moisture level
10	CWHWS wheat at 20% moisture level
11	CWRS wheat at 12% moisture level
12	CWRS wheat at 14% moisture level
13	CWRS wheat at 16% moisture level
14	CWRS wheat at 18% moisture level
15	CWRS wheat at 20% moisture level
16	CWRW wheat at 12% moisture level
17	CWRW wheat at 14% moisture level
18	CWRW wheat at 16% moisture level
19	CWRW wheat at 18% moisture level
20	CWRW wheat at 20% moisture level
21	CWSWS wheat at 12% moisture level
22	CWSWS wheat at 14% moisture level
23	CWSWS wheat at 16% moisture level
24	CWSWS wheat at 18% moisture level
25	CWSWS wheat at 20% moisture level

Table A.5 Confusion matrix for the 75 NIR absorbance feature LDA model for five wheat classes each at five various moisture levels (n = 100 per class per moisture level)

Class (To) → / (From) ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	99	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	98	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	94	0	0	0	0	5	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	86	0	0	0	0	9	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	98	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	86	1	0	0	12	0	0	0	0	0	1	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	98	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	2	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	2	1	0	0	0	0	96	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	4	0	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	1	0	0	0	0	98	1	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	1	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	2	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	98	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	98	1	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	1
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	98	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

Table A.6 Confusion matrix for the 75 NIR absorbance feature QDA model for five wheat classes each at five various moisture levels (n = 100 per class per moisture level)

Class (To) → / (From) ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	99	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	99	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	99	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	85	0	0	0	1	4	4	0	0	1	3	1	0	0	0	0	0	0	0	0	1	0
5	0	0	0	0	98	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	77	0	0	0	0	22	0	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	89	0	0	0	4	0	2	0	0	0	3	0	0	0	0	1	1	0	0
8	0	0	0	3	0	0	0	86	0	0	0	0	9	0	0	0	0	0	0	0	0	0	1	1	0
9	0	0	0	6	0	0	0	0	89	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	10	0	0	0	0	2	77	0	0	0	3	7	0	0	0	0	0	0	0	0	1	0
11	1	1	0	0	0	2	0	0	0	0	93	0	0	0	0	1	2	0	0	0	0	0	0	0	0
12	0	2	0	0	0	0	0	0	0	0	0	92	6	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	2	0	0	0	2	0	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	5	0	0	0	1	0	0	0	0	3	91	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	1	1	0	0	0	2	96	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	85	11	0	0	0	1	1	0	0	0
17	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	1	96	0	0	0	1	0	0	0	0
18	0	1	0	2	0	0	0	0	0	0	0	0	2	0	0	0	2	90	1	0	0	0	0	2	0
19	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	97	0	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95	0	0	0	0	5
21	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	4	0	0	0	0	92	1	0	0	0
22	0	5	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	93	0	0	0
23	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	95	2	0
24	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	95	1
25	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99

Table A.7 Confusion matrix for the 51 NIR absorbance feature LDA model for five wheat classes each at five various moisture levels (n = 100 per class per moisture level)

Class (To) → / (From) ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	98	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	98	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	86	0	0	0	0	12	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	71	0	0	0	0	23	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	92	2	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	98	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	4	0	0	0	1	94	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	4	0	0	0	0	3	92	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	6	0	0	0	0	94	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	1	0	0	0	0	98	1	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	1	0	0	0	0	1	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	1	0	0	0	0	2	0	0	0	1	96	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	3	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	97	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	1	0	0	0	0	1
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	99	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	97	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100

Table A.8 Confusion matrix for the 51 NIR absorbance feature QDA model for five wheat classes each at five various moisture levels (n = 100 per class per moisture level)

Class (To) → / (From) ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	87	0	0	0	0	6	3	0	0	1	3	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	95	0	0	0	0	1	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	85	0	0	0	0	13	0	0	0	0	2	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	92	0	0	0	2	0	2	0	0	0	3	0	0	0	0	0	1	0	0
8	0	0	0	1	0	0	0	93	1	0	0	0	4	0	0	0	0	1	0	0	0	0	0	0	0
9	0	0	0	13	0	0	0	0	85	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	10	0	0	0	0	1	81	0	0	0	1	7	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	2	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	2	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	2	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	1	0	0	0	0	0	1	0	0	0	98	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	1	0	99	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	92	5	0	0	0	2	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	99	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	97	2	0	0	0	0	0	0
19	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	0	0	0	0	2
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	99	0	0	0	0
22	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	96	0	0	0
23	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	1	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	2
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	99

Table A.9 Confusion matrix of top seven NIR absorbance feature LDA model for five wheat classes each at five various moisture levels (n = 100 per class per moisture level)

Class (To) → / (From) ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	96	0	0	0	0	1	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
2	0	93	0	0	0	1	1	0	0	0	0	4	0	0	0	0	0	0	1	0	0	0	0	0	0
3	0	0	89	0	0	0	0	10	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
4	0	0	0	78	0	0	0	0	9	0	0	0	2	11	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	54	0	0	0	0	39	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	96	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	5	87	1	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	1	0	0	0	0	91	3	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	13	0	0	0	1	81	0	0	0	1	4	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	2	2	0	0	0	4	76	0	0	0	1	11	0	0	0	0	4	0	0	0	0	0
11	0	0	0	0	0	16	1	0	0	0	82	1	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	1	1	0	0	0	1	0	0	0	1	93	3	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	3	0	0	0	3	3	0	0	0	90	1	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	14	0	0	0	0	1	0	0	0	1	83	1	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	3	0	0	0	0	0	0	0	0	5	91	0	0	0	0	1	0	0	0	0	0
16	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	91	3	0	0	0	0	0	0	0	0
17	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	97	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	97	1	0	0	0	1	0	0
19	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	98	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	97	0	0	0	0	3
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	93	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	97	1	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	5	91	3
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	98

Table A.10 Confusion matrix of top seven NIR absorbance feature QDA model for five wheat classes each at five various moisture levels (n = 100 per class per moisture level)

Class (To) → / (From) ↓	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	99	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	77	0	0	0	0	10	2	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	85	0	0	0	0	8	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	92	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	92	1	0	0	5	0	0	0	0	0	2	0	0	0	0	0	0	0	0
8	0	0	0	1	0	0	0	93	2	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	12	0	0	0	1	80	0	0	0	5	2	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	2	21	0	0	0	1	73	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
11	0	0	0	0	0	11	0	0	0	0	89	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	2	0	0	0	0	0	0	0	0	1	94	3	0	0	0	0	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	2	2	0	0	0	94	2	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	10	0	0	0	0	1	2	0	0	1	84	2	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	2	0	0	0	0	0	0	0	0	1	97	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	98	1	0	0	0	1	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	96	0	0	0	0	2	0	0	0
18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	98	1	0	0	0	0	0	0
19	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	93	0	0	0	0	6
21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	98	0	0	0
23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100	0	0
24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	93	5
25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1	96

Table A.11 Predicted protein contents of wheat classes using PLSR model (75 input features).

Sample No.	CWHWS	CWES	CWRS	CWRW	CWSWS
1	16.87	16.94	20.25	13.62	14.26
2	17.79	16.81	19.17	14.52	14.51
3	17.49	16.11	19.35	13.21	15.03
4	17.12	16.74	18.00	13.75	15.17
5	16.59	16.80	18.32	14.36	15.02
6	16.96	16.88	19.98	14.40	14.38
7	16.93	16.10	18.92	13.75	13.48
8	17.14	16.44	19.23	14.05	15.00
9	17.36	16.51	19.08	13.55	14.55
10	17.77	16.00	18.85	14.16	14.88
11	16.34	17.37	19.16	13.55	14.26
12	17.53	16.14	18.67	14.43	14.58
13	17.13	17.03	18.23	14.33	13.72
14	16.67	17.08	18.29	15.16	14.40
15	17.22	16.70	18.35	14.05	14.34
16	17.92	17.13	18.42	14.37	14.45
17	17.03	16.87	18.62	14.42	14.48
18	16.62	17.61	19.00	13.21	13.01
19	17.23	16.69	18.28	14.19	13.41
20	16.98	17.86	17.21	13.72	14.28
21	17.69	17.52	18.89	14.30	14.24
22	17.37	15.78	18.96	14.29	14.47
23	17.54	15.57	18.68	14.49	14.21
24	18.22	16.09	18.43	14.49	14.88
25	17.80	16.29	17.75	14.31	13.66
26	17.69	17.09	18.25	14.34	14.38
27	17.51	15.33	18.64	13.76	14.27
28	18.32	15.93	18.62	15.19	14.36
29	17.97	16.52	18.98	14.07	13.78
30	17.23	16.47	18.20	14.68	13.80
31	17.06	16.88	18.94	14.53	14.35
32	17.99	16.58	19.08	15.22	14.59
33	17.01	18.07	19.59	15.09	14.27
34	17.24	18.17	18.20	14.61	14.63
35	17.66	18.07	19.17	14.38	14.21
36	17.11	16.80	20.03	14.85	14.65
37	17.38	18.65	19.22	15.15	14.89
38	17.28	16.71	19.91	14.18	13.71
39	17.76	17.14	19.42	14.70	14.49
40	17.26	17.15	19.51	14.82	14.54
41	17.52	16.70	19.60	14.47	13.42
42	17.40	16.60	18.74	15.50	13.35
43	18.24	15.27	18.59	14.93	13.28
44	17.93	15.00	18.52	14.32	13.32
45	17.16	15.38	18.17	14.74	15.35
46	17.11	14.86	18.52	14.59	14.20
47	18.23	16.66	19.05	14.61	14.57
48	17.83	15.76	18.58	14.09	13.91
49	17.86	16.28	18.05	14.54	13.71
50	17.26	15.53	18.54	14.40	13.77

Table A.12 Predicted protein contents of wheat classes using PLSR model (51 input features).

Sample No.	CWHWS	CWES	CWRS	CWRW	CWSWS
1	16.60	17.24	19.99	13.73	13.87
2	17.54	17.18	19.53	14.56	14.26
3	17.15	16.52	19.85	13.59	14.12
4	16.61	17.06	18.98	14.12	14.58
5	16.66	17.16	18.53	14.40	14.41
6	16.98	17.18	20.25	14.62	13.79
7	16.95	16.54	19.29	13.55	13.03
8	16.93	16.89	19.59	14.26	14.43
9	16.97	16.79	19.44	13.73	13.98
10	17.13	16.26	19.17	14.25	14.13
11	16.57	17.57	19.38	13.41	13.65
12	17.08	16.40	18.81	14.30	13.80
13	17.01	17.49	18.35	14.09	13.24
14	16.21	17.18	18.48	14.93	14.03
15	17.51	17.20	18.56	13.98	13.63
16	17.83	17.49	18.60	14.04	13.96
17	17.00	17.01	18.64	14.45	13.80
18	16.93	17.76	19.21	12.37	12.44
19	17.15	16.82	18.44	13.30	13.98
20	17.14	18.15	17.58	13.48	14.01
21	17.14	17.53	18.61	14.67	14.12
22	17.32	15.85	18.70	14.72	14.43
23	17.12	15.71	18.98	14.90	14.09
24	17.87	16.17	18.68	14.89	14.93
25	17.84	16.30	17.77	14.93	13.49
26	17.52	16.90	18.01	14.81	14.29
27	17.58	15.43	18.57	14.13	14.32
28	18.08	16.10	18.47	15.23	14.28
29	17.98	16.38	19.02	14.65	13.45
30	17.11	16.63	18.05	15.19	13.61
31	16.75	16.93	18.42	14.79	14.40
32	17.57	17.29	18.58	15.11	14.56
33	16.74	17.84	19.09	15.29	14.14
34	16.91	18.34	17.83	14.62	14.81
35	17.45	18.71	19.05	14.73	14.52
36	16.70	16.71	19.80	15.03	15.02
37	16.81	18.86	18.65	14.88	15.27
38	16.97	17.46	19.16	14.15	14.04
39	17.23	17.20	19.05	14.82	14.85
40	16.83	17.42	18.99	14.66	14.87
41	17.55	16.07	19.30	14.64	14.27
42	17.10	16.34	18.45	15.55	14.11
43	17.95	15.15	18.45	14.85	14.18
44	17.70	15.21	18.35	14.42	14.13
45	17.06	15.31	17.93	15.00	15.89
46	17.47	14.74	18.42	14.54	14.93
47	17.93	15.92	19.06	14.85	15.47
48	17.89	15.23	18.69	14.11	14.69
49	17.69	15.94	18.07	14.67	14.42
50	17.17	15.31	18.70	14.40	14.57

Table A.13 Predicted oil contents of wheat classes using PLSR model (75 input features).

Sample No.	CWHWS	CWES	CWRS	CWRW	CWSWS
1	1.52	1.57	1.40	1.81	1.79
2	1.48	1.52	1.41	1.73	1.76
3	1.50	1.55	1.28	1.78	1.77
4	1.54	1.53	1.36	1.78	1.71
5	1.63	1.60	1.40	1.88	1.75
6	1.59	1.58	1.30	1.74	1.82
7	1.57	1.53	1.36	1.84	1.84
8	1.53	1.56	1.36	1.77	1.77
9	1.51	1.55	1.38	1.81	1.72
10	1.49	1.53	1.35	1.79	1.78
11	1.62	1.44	1.35	1.82	1.71
12	1.54	1.54	1.34	1.80	1.76
13	1.55	1.44	1.42	1.83	1.78
14	1.56	1.50	1.38	1.74	1.80
15	1.47	1.46	1.36	1.75	1.69
16	1.60	1.40	1.40	1.80	1.77
17	1.59	1.49	1.39	1.70	1.75
18	1.56	1.46	1.40	1.79	1.88
19	1.59	1.48	1.39	1.88	1.66
20	1.61	1.39	1.42	1.76	1.72
21	1.53	1.47	1.50	1.71	1.55
22	1.53	1.52	1.50	1.72	1.57
23	1.55	1.52	1.45	1.73	1.60
24	1.45	1.49	1.41	1.73	1.58
25	1.53	1.46	1.53	1.70	1.66
26	1.47	1.37	1.43	1.73	1.57
27	1.58	1.49	1.48	1.81	1.59
28	1.46	1.44	1.45	1.69	1.56
29	1.53	1.50	1.48	1.75	1.63
30	1.62	1.40	1.44	1.71	1.63
31	1.69	1.38	1.43	1.75	1.66
32	1.59	1.42	1.40	1.72	1.66
33	1.81	1.36	1.31	1.69	1.67
34	1.68	1.38	1.43	1.73	1.61
35	1.60	1.29	1.40	1.76	1.69
36	1.67	1.42	1.26	1.74	1.63
37	1.69	1.33	1.44	1.73	1.60
38	1.63	1.38	1.32	1.79	1.67
39	1.60	1.43	1.30	1.72	1.65
40	1.70	1.34	1.29	1.80	1.63
41	1.51	1.51	1.38	1.76	1.74
42	1.56	1.49	1.42	1.66	1.75
43	1.46	1.60	1.43	1.68	1.80
44	1.49	1.57	1.45	1.73	1.71
45	1.56	1.60	1.46	1.72	1.71
46	1.49	1.56	1.40	1.74	1.75
47	1.52	1.45	1.38	1.67	1.68
48	1.47	1.57	1.38	1.75	1.71
49	1.53	1.59	1.46	1.73	1.78
50	1.56	1.60	1.40	1.80	1.74

Table A.14 Predicted oil contents of wheat classes using PLSR model (51 input features).

Sample No.	CWHWS	CWES	CWRS	CWRW	CWSWS
1	1.52	1.56	1.37	1.80	1.73
2	1.43	1.53	1.39	1.71	1.68
3	1.48	1.58	1.27	1.79	1.73
4	1.48	1.54	1.32	1.73	1.64
5	1.60	1.58	1.38	1.74	1.70
6	1.57	1.60	1.28	1.71	1.77
7	1.54	1.55	1.33	1.82	1.78
8	1.53	1.57	1.32	1.71	1.71
9	1.51	1.57	1.35	1.79	1.70
10	1.51	1.56	1.33	1.77	1.73
11	1.59	1.47	1.36	1.79	1.77
12	1.49	1.55	1.37	1.74	1.79
13	1.47	1.45	1.43	1.79	1.82
14	1.51	1.50	1.42	1.68	1.81
15	1.38	1.46	1.37	1.71	1.78
16	1.54	1.44	1.43	1.71	1.82
17	1.56	1.52	1.43	1.67	1.80
18	1.54	1.47	1.44	1.78	1.92
19	1.54	1.49	1.42	1.70	1.73
20	1.56	1.39	1.46	1.68	1.80
21	1.56	1.42	1.48	1.70	1.65
22	1.57	1.48	1.49	1.70	1.64
23	1.58	1.48	1.47	1.70	1.66
24	1.49	1.46	1.46	1.69	1.65
25	1.56	1.45	1.51	1.68	1.72
26	1.49	1.35	1.47	1.70	1.65
27	1.60	1.50	1.49	1.79	1.64
28	1.47	1.44	1.48	1.68	1.63
29	1.52	1.46	1.43	1.68	1.72
30	1.63	1.38	1.46	1.68	1.70
31	1.64	1.42	1.36	1.74	1.78
32	1.55	1.45	1.35	1.70	1.75
33	1.64	1.36	1.29	1.72	1.78
34	1.60	1.39	1.39	1.75	1.66
35	1.54	1.32	1.33	1.74	1.70
36	1.64	1.35	1.23	1.73	1.66
37	1.63	1.32	1.41	1.71	1.64
38	1.58	1.40	1.24	1.78	1.70
39	1.59	1.45	1.25	1.71	1.69
40	1.64	1.41	1.24	1.77	1.67
41	1.55	1.55	1.40	1.72	1.75
42	1.54	1.54	1.39	1.66	1.76
43	1.47	1.61	1.44	1.68	1.78
44	1.54	1.58	1.47	1.73	1.71
45	1.57	1.62	1.49	1.71	1.73
46	1.49	1.57	1.45	1.72	1.77
47	1.53	1.49	1.42	1.67	1.71
48	1.46	1.58	1.40	1.74	1.75
49	1.52	1.60	1.45	1.72	1.78
50	1.57	1.61	1.41	1.78	1.75