

**DETECTION OF INSECT AND FUNGAL DAMAGE  
AND INCIDENCE OF SPROUTING IN STORED  
WHEAT USING NEAR-INFRARED  
HYPER SPECTRAL AND DIGITAL COLOR IMAGING**

**A Thesis**

**Submitted to the Faculty of Graduate Studies**

**The University of Manitoba**

**in partial fulfilment of the requirements for the degree of**

***Doctor of Philosophy***

**By**

**Chandra Bhan Singh**

**Department of Biosystems Engineering**

**University of Manitoba**

**Winnipeg, Canada**

**R3T 5V6**

**© August 2009**

**THE UNIVERSITY OF MANITOBA  
FACULTY OF GRADUATE STUDIES  
COPYRIGHT PERMISSION PAGE**

**DETECTION OF INSECT AND FUNGAL DAMAGE AND  
INCIDENCE OF SPROUTING IN STORED WHEAT USING NEAR-  
INFRARED HYPERSPECTRAL AND DIGITAL COLOR IMAGING**

**By**

**CHANDRA BHAN SINGH**

**A Thesis/Practicum submitted to the Faculty of Graduate Studies of  
The University of Manitoba in partial fulfilment of the requirements for the degree  
of**

**Doctor of Philosophy**

**CHANDRA BHAN SINGH © 2009**

**Permission has been granted to the library of The University of Manitoba to lend or sell copies of this thesis/practicum, to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film, and to University microfilm Inc. to publish an abstract of this thesis/practicum.**

**The author reserves other publication rights, and neither this thesis/practicum nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.**

## ABSTRACT

Canada has a reputation of exporting high quality grain globally. Wheat grain quality is defined by several parameters, of which insect and fungal damage and sprouting are considered important degrading factors. There is an increasing demand among grain buyers and consumers towards zero-tolerance to contamination in grain and processed grain products. At present, Canadian wheat is inspected and graded manually by Canadian Grain Commission (CGC) inspectors at grain handling facilities or in the CGC laboratories. Visual inspection methods are time consuming, less efficient, subjective, and require experienced personnel. Therefore, an alternative, rapid, objective, accurate, and cost effective technique is needed for grain quality monitoring in real-time which can potentially assist or replace the manual inspection process.

Insect-damaged wheat samples by the species of rice weevil (*Sitophilus oryzae*), lesser grain borer (*Rhyzopertha dominica*), rusty grain beetle (*Cryptolestes ferrugineus*), and red flour beetle (*Tribolium castaneum*); fungal-damaged wheat samples by the species of storage fungi namely *Penicillium spp.*, *Aspergillus glaucus* group, and *Aspergillus niger*; and artificially sprouted and midge-damaged wheat kernels were obtained from the Cereal Research Centre (CRC), Agriculture and Agri-Food Canada, Winnipeg, Canada. Healthy and damaged wheat kernels were imaged using a long-wave near-infrared (LWNIR) and a short-wave near-infrared (SWNIR) hyperspectral imaging system and an area-scan color camera.

The hyperspectral data from both LWNIR and SWNIR were analyzed using multivariate image analysis (MVI) after applying pre-processing and calibration techniques. The wavelengths corresponding to the highest principal components (PC)

factor loadings were considered to be significant. Up to six statistical features (maximum, minimum, mean, median, standard deviation, and variance) and histogram features from the significant wavelengths images were extracted and given as input to various classifiers. From the color images of individual kernels, 230 features (123 colour, 56 textural, and 51 morphological features) were extracted and the top 10 features from 123 color features and combined 230 features were selected and given as input to various classifiers. These top features were also combined with the SWNIR features to develop classification models. Linear, quadratic, and Mahalanobis discriminant analysis classifiers were used to develop supervised classification algorithms. Back propagation neural network (BPNN) classifiers were also investigated for their performance.

Long-wave NIR hyperspectral imaging gave high accuracy in classifying healthy and damaged wheat kernels and detected 85.0-100.0% healthy and insect-damaged kernels, 95-100.0% healthy and fungal-infected kernels, and 85.0-100.0% healthy and sprouted kernels in two-way classification. The SWNIR hypersepcrtal imaging gave high damage detection rate but caused high false positive errors of up to 28.0% and misclassified healthy kernels as damaged. In some cases, the top features from color images also resulted in high false positive errors. Combined features from color and SWNIR images gave very high classification accuracy and detected 91.0-100.0% healthy and insect damaged, 99.0-100.0% healthy and fungal damaged, and 95.3- 99.3% healthy and sprouted (midge-damaged) wheat kernels. Classification results of statistical discriminant classifiers were better than the BPNN classifiers.

## ACKNOWLEDGEMENTS

With profound indebtedness, I wish to acknowledge my deep sense of gratitude to Dr. Digvir Singh Jayas for giving me an opportunity to pursue my doctoral research under his supervision, his excellent guidance, and support.

I am very grateful to Dr. Jitendra Paliwal for co-supervising me and for his valuable suggestions, encouragement, and support. My due thanks is also to Dr. N.D.G. White for being in my research committee, his guidance, and support in sample preparation at the Cereal Research Centre lab. I am also thankful to my Advisory Committee members Dr. Reza Fazel and Dr. Gabriel Thomas.

I thank the Biosystems Engineering Department staff members and technicians for their support and help. I also thank Ms. Parneet Mavi, Mr. Richard Dyrkacz, and Ms. Dipali Narvankar for their help in data collection. I am very thankful to Dr. Prabal Ghosh for his help and support from the moment I landed in Winnipeg and throughout my PhD program.

I wish to acknowledge the support and cooperation I received from my fellow researchers and thank Dr. Ruplal Choudhary, Dr. Hao Zhang, Dr Manickavasagan Annamalai, Dr. Vadivambal Rajagopal, Dr. Sabyasachi Mishra, Mr. Aravind Mohan Lokhamoorthi, Ms. Sathya Gunashekran, Mr. Anandakumar Palanichamy, Mr. Mahesh Sivakumar, Mr. Suresh Neethirajan, Mr. Jun Wu, Mr. Feng Wang, Ms. Ke Sun, Ms. Sujala Balaji, Ms. Hai Yan Li, Ms. Gayathri Pitchai, Ms. Nithya Udyakumar, Ms. Deepali Narvankar, Mr. Chelladurai Vellaichamy, Mr. Rajarammanna Ramachandran, Mr. Ganesan Ramalingam, and Mr. Senthil Tiruppathi.

I also wish to express my gratitude to my parents, brothers, and sister for their support to me. I thank my little daughter Shriya and my beloved wife Bandana for her moral support and being my strength in all situations.

# TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	xv
LIST OF SYMBOLS	xvii
1. INTRODUCTION	1
2. THEORETICAL BACKGROUND	8
2.1 Electromagnetic Spectrum	8
2.1.1 Mechanical models	10
2.1.1.1 Harmonic model	10
2.1.1.2 Quantum mechanical theory	11
2.1.1.3 Anharmonic model	13
2.2 NIR Spectroscopy	14
2.3 Color Imaging	15
2.4 Hyperspectral Imaging	16
2.5 Instrumentation	16
2.5.1 Color imaging system	16
2.5.1.1 Lighting source	16
2.5.1.2 Sensor	17
2.5.1.3 Communication interface	18
2.5.2 Hyperspectral imaging system	18

2.5.2.1 Detectors	19
2.5.2.2 Wavelength filtering devices	21
2.5.2.3 Illumination sources	23
2.5.2.4 Integration of software and hardware	23
2.6 Data Analysis	24
3. SYNTHESIS OF LITERATURE	26
3.1 Detection of Insect, Fungal, and Sprout Damage	26
3.1.1 Insect damage	26
3.1.2 Fungal damage	27
3.1.3 Sprout damage	29
3.2 Applications of NIR Spectroscopy	30
3.3 Applications of Digital Color Imaging	38
3.4 Applications of NIR Hypespectral Imaging	42
3.5 Summary of Literature Reviewed	51
4. MATERIALS AND METHODS	53
4.1 Sample Preparation	53
4.1.1 Insect infested kernels	53
4.1.2 Fungal-infected kernels	53
4.1.3 Artificially-sprouted and midge-damaged kernels	54
4.2 Imaging Systems	55
4.2.1 Long-wave NIR imaging system	55
4.2.1.1 Image acquisition	56
4.2.1.2 Data analysis	57

4.2.2 Short-wave NIR imaging system	59
4.2.2.1 Image acquisition	62
4.2.2.2 Data analysis	63
4.2.3 Color imaging system	63
4.2.3.1 Image acquisition	64
4.2.3.2 Data analysis	64
4.3 Model Development	66
5. RESULTS AND DISCUSSION	68
5.1 Insect Damage Detection	68
5.1.1 Detection of insect-damaged kernels using LWNIR hyperspectral imaging	68
5.1.2 Detection of insect-damaged kernels using SWNIR hyperspectral imaging	80
5.1.3 Detection of insect-damaged kernels using color imaging	83
5.1.4 Detection of insect-damaged kernels using color and SWNIR imaging	87
5.1.5 Classification of insect-damaged kernels using BPNN	88
5.2 Fungal Damage Detection	91
5.2.1 Fungal detection using LWNIR hyperspectral imaging	91
5.2.2 Detection of fungal-infected kernels using SWNIR hyperspectral imaging	98
5.2.3 Detection of fungal-infected kernels using color imaging	104
5.2.4 Detection of fungal-infected kernels using color and SWNIR	107

imaging	
5.3 Detection of Sprout Damage	108
5.3.1 Detection of sprouted and midge-damaged kernels using LWNIR	108
hyperspectral imaging	
5.3.2 Detection of midge-damaged kernels using SWNIR hyperspectral	117
imaging	
5.3.3 Detection of midge-damaged kernels using color imaging	120
5.3.4 Detection of midge-damaged kernels using color and SWNIR	123
imaging	
6. CONCLUSIONS AND RECOMENDATIONS	127
7. REFERENCES	130
8. APENDIX	150

## LIST OF TABLES

<b>Number</b>	<b>Title</b>	<b>Page</b>
Table 5.1	Binary classification (20% test set) using statistical features (1101.69 and 1305.05 nm wavelengths) for healthy wheat kernels and those damaged by various insect species	74
Table 5.2	Binary classification (20% test set) using statistical features (1101.69, 1305.05, and 1447.46 nm wavelengths) for healthy wheat kernels and those damaged by various insect species	74
Table 5.3	Binary classification (20% test set) using histogram features (1101.69 nm wavelength) for healthy wheat kernels and those damaged by various insect species	77
Table 5.4	Binary classification (20% test set) using statistical (1101.69 and 1305.05 nm wavelengths) and histogram features (1101.69 nm wavelength) for healthy wheat kernels and those damaged by various insect species	77
Table 5.5	Two-way classification (20% test set) using statistical features (1101.69 and 1305.05 nm wavelengths) for healthy wheat kernels and those damaged by various insect species	78
Table 5.6	Two-way classification (20% test set) using statistical features (1101.69, 1305.05, and 1447.46 nm wavelengths) for healthy wheat kernels and those damaged by various insect species	78

Table 5.7	Two-way classification (20% tests set) using histogram features (1101.69 nm wavelength) for healthy wheat kernels and those damaged by various insect species	79
Table 5.8	Two-way classification (20% tests set) using statistical (1101.69 and 1305.05 nm wavelengths) and histogram features (1101.69 nm wavelength) for healthy wheat kernels and those damaged by various insect species	79
Table 5.9	Two-way classification of healthy and insect-damaged wheat kernels by four insect species using SWNIR hyperspectral image features	84
Table 5.10	Top 10 features selected from 123 color features of healthy and insect-damaged wheat kernels	85
Table 5.11	Two-way classification of healthy and insect-damaged wheat kernels using top 10 color features from color images	85
Table 5.12	Top 10 features selected from 230 morphological, color, and textural features of healthy and insect-damaged wheat kernels	86
Table 5.13	Two-way classification of healthy and insect-damaged kernels using top 10 features from 230 morphological, color, and textural features extracted from color images	86
Table 5.14	Two-way classification of healthy and insect-damaged wheat kernels using combined top 9 features from 123 color features and SWNIR image features	89
Table 5.15	Two-way classification of healthy and insect-damaged kernels using combined top 10 features from 230 morphological, color, and textural features and SWNIR image features	89

Table 5.16	Two-way classification of healthy and insect-damaged wheat kernels by back-propagation neural network (BPNN) classifier	90
Table 5.17	Pair-wise discrimination of healthy and fungal-infected wheat kernels by k-means classifier using highest, lowest, and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths	94
Table 5.18	Pair-wise discrimination of healthy and fungal-infected wheat kernels by discriminant analysis using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths	94
Table 5.19	Two-way discriminant classification analysis to separate fungal-infected kernels from healthy kernels using highest, lowest, and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths	96
Table 5.20	Two-way classification analysis using test set of 60 mixed kernels (48 healthy and 12 infected kernels) and highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths	96
Table 5.21	Pair-wise discrimination between kernels infected with various types of fungi by discriminant classifiers using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths	99
Table 5.22	Four-class classification (Linear discriminant analysis) using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths	99

Table 5.23	Four-class classification (Quadratic discriminant analysis) using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths	100
Table 5.24	Four-class classification (Mahalanobis discriminant analysis) using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths	100
Table 5.25	Two-way classification of healthy and fungal-infected wheat kernels using SWNIR hyperspectral image features	105
Table 5.26	Top 10 features selected from 123 color features of healthy and fungal-infected wheat kernel	105
Table 5.27	Two-way classification of healthy and fungal-infected wheat kernels using top 8 color features from 123 color features	106
Table 5.28	Top 10 features selected from 230 morphological, color, and textural features of healthy and fungal-infected wheat kernels	106
Table 5.29	Two-way classification of healthy and fungal-infected kernels using top 10 features from 230 morphological, color, and textural features	109
Table 5.30	Two-way classification of healthy and fungal-infected kernels using combined top 8 color features from 130 color features and SWNIR hyperspectral image features	109
Table 5.31	Two-way classification of healthy and fungal-infected kernels using combined top 10 features from 230 morphological, color, and textural and SWNIR image features	109

Table 5.32	Classification of sprout-damaged wheat kernels using 300 healthy and 300 sprouted kernels (80% training and 20% test sets) (LWNIR)	115
Table 5.33	Classification of midge-damaged wheat kernels (crease-down) using 300 healthy and 300 midge-damaged kernels from each of five locations (80% training and 20% test sets) (LWNIR)	116
Table 5.34	Classification of midge-damaged wheat kernels (crease-up) using 300 healthy and 300 midge-damaged kernels from each of five locations (80% training and 20% test sets) (LWNIR)	116
Table 5.35	Two-way classification of healthy and midge-damaged wheat kernels using SWNIR hyperspectral image features	120
Table 5.36	Top 10 features selected from 123 color features of healthy and midge-damaged wheat kernels	122
Table 5.37	Two-way classification of healthy and midge-damaged kernels using top 10 color features from 123 color features extracted from color images of wheat from different locations in western Canada	122
Table 5.38	Top 10 features selected from 230 morphological, color, and textural features of midge-damaged wheat kernels	124
Table 5.39	Two-way classification of healthy and midge-damaged wheat kernels using top 10 features from 230 morphological, color, and textural features extracted from color images of wheat from different locations in western Canada	124

Table 5.40	Two-way classification of healthy and midge-damaged wheat kernels using combined top 10 color features from 123 color features and SWNIR hyperspectral image features from different locations in western Canada	126
Table 5.41	Two-way classification of healthy and midge-damaged kernels using combined top 10 features from 230 morphological, color and textural features and SWNIR hyperspectral image features from different locations in western Canada	126
Table A.1	Applications of NIR spectroscopy in grain quality analysis	151
Table A.2	Applications of imaging processing in grain quality analysis	152
Table A.3	Reported applications of hyperspectral imaging	153
Table A.4	Four-way classification of healthy and fungal-infected kernels by LDA using SWNIR image features	154
Table A.5	Four-way classification of healthy and fungal-infected kernels by QDA using SWNIR image features	154
Table A.6	Four-way classification of healthy and fungal-infected kernels by Mahalanobis classifier using SWNIR image features	154

## LIST OF FIGURES

<b>Number</b>	<b>Title</b>	<b>Page</b>
Fig. 2.1	Electromagnetic spectrum	9
Fig. 2.2	Diatomic harmonic (A) and anharmonic (B) models and their potential energy	12
Fig 4.1	Long-wave near-infrared hyperspectral imaging system	60
Fig. 4.2	(a) NIR image and (b) labelled kernels in image (a) separated from the background	60
Fig 4.3	Short-wave near-infrared hyperspectral imaging system	61
Fig 4.4	Color imaging system	65
Fig. 5.1	First principal component (PC) mean factor loadings of healthy and insect-damaged kernels (long-wave near-infrared)	69
Fig. 5.2	Second principal component (PC) mean factor loadings (absolute values) of healthy and insect-damaged kernels (long-wave near-infrared)	70
Fig. 5.3	Images of healthy and insect-damaged kernels (NIR images (1305.05 nm) on left and first PC score images on right)	72
Fig. 5.4	Short-wave near-infrared absorbance mean spectra of healthy and insect-damaged wheat kernels	81
Fig. 5.5	First principal component (PC) mean factor loadings of healthy and insect-damaged wheat kernels (short-wave near-infrared)	82
Fig. 5.6	First principal component (PC) loadings of healthy and fungal-infected wheat kernels (long-wave near-infrared)	92
Fig. 5.7	Short-wave near-infrared absorbance mean spectra of healthy and	101

	fungus-infected wheat kernels	
Fig. 5.8	First principal component (PC) mean factor loadings of healthy and fungus-infected wheat kernels (short-wave near-infrared)	102
Fig. 5.9	First principal component (PC) scores images of healthy, sprouted, and midge-damaged wheat kernels (long-wave near-infrared)	111
Fig. 5.10	PC loadings of first principal component (PC) of healthy and midge-damaged wheat kernels for different locations in western Canada (long-wave near-infrared)	113
Fig. 5.11	Long-wave near-infrared mean reflectance spectra of healthy and midge-damaged wheat kernels for different locations in western Canada	114
Fig. 5.12	Short-wave near-infrared mean reflectance spectra of healthy and midge-damaged wheat kernels for different locations in western Canada	118
Fig. 5.13	First principal component (PC) mean factor loadings of healthy and midge-damaged wheat kernels (short-wave near-infrared) for different locations in western Canada	119

## LIST OF SYMBOLS

ANN	Artificial Neural Network
AOTF	Acousto-Optical Tunable Filter
API	Active Pharmaceutical Ingredient
ASCC	Average Squared Canonical Correlation
BPNN	Back Propagation Neural Network
CCD	Charge-coupled Device
CCIR	Consultative Committee on International Radio
CGC	Canadian Grain Commission
CLS	Classical Least Square
CMOS	Complementary Metal Oxide Semiconductor
CWAD	Canada Western Amber Durum
CWRS	Canada Western Red Spring
CWSWS	Canada Western Soft White Spring
DA	Discriminant Analysis
DHV	Dark Hard Vitreous
DT	Detreding
ETF	Electronically Tunable Filters
FDA	Fisher's Discriminant Analysis
FOV	Field of View
FPA	Focal Plane Arrays
FT	Fourier Transform
FWHM	Full Width at Half Maximum

GA	Genetic Algorithm
GLCM	Gray Level Co-occurrence Matrix
GLRM	Gray Level Run Length Matrix
HDWW	Hard White Winter
HPLC	High Performance Liquid Chromatography
HRS	Hard Red Spring
HRW	Hard Red Winter
ICA	Independent Component Analysis
IR	Infrared
KNN	K-Nearest Neighbors
LCTF	Liquid Crystal Tunable Filter
LDA	Liner Disrciminant Analysis
LED	Light Emitting Diodes
LS-SVM	Least-squares Support Vector Machines
LWNIR	Long-wave NIR
MC	Mean Centering
MDS	Multidimensional Scaling
MLR	Multiple Linear Regression
MNF	Minimum Noise Fraction
MRI	Magnetic Resonance Imaging
MSC	Multiple Scatter Correction
MVI	Multivariate Image Analysis
NDHV	Non-dark hard vitreous PLS;

NIR	Near-infrared
NTSC	National Television System Committee
PAL	Phase Alternating Line
PCA	Principal Component Analysis
PCI	Peripheral Component Interconnect
PDA	Penalized Discriminant Analysis
PLS	Partial Least Square
PNN	Probabilistic Neural Network
QDA	Quadratic Discriminant Analysis
RF	Radio Frequencies
RGB	Red, Green, and Blue
ROI	Region of Interest
SEC	Standard Error of Calibration
SECV	Standard Error of Cross-Validation
SE-HPLC	Size Exclusive-High Performance Liquid Chromatography
SEP	Standard Error of Prediction
SG	Savitzky-Golay
SMV	Soybean Mosaic Virus
SNV	Standard Normal Variate
SRW	Soft Red Winter
SVM	Support Vector Machine
SWNIR	Short-wave NIR
UV	Ultraviolet

VIS	Visible
WT	Wavelet Transform

## 1. INTRODUCTION

Canada produces 52.7 Mt (million tonnes) of cereal grains annually (FAOSTAT, 2009), of which 60-70% is exported. A total of 21 crops are registered as grains in Canada (CGC, 2006) including cereals (wheat, barley, corn, oats, and rye), oil seeds (canola, soybeans, safflower, sunflower, flaxseed), and pulses (peas, beans, chickpeas, and lentils). Canada is the sixth largest wheat-producing country, ranks among the largest exporters in the world with 25.3 Mt average annual production and 18.5 Mt annual export (FAO, 2006), and has a reputation of exporting high quality grain in the world grain market. Usually farmers store grains in on-farm free standing bins after the harvest until needed to be transferred to the primary elevators (grain handling facility) by trucks. Primary elevators receive, weigh the grain, determine the dockage, assign a grade based on rapid visual inspection, and store the grain before it is sold and transferred to the terminal elevators, mostly by rail cars. Terminal elevators receive, clean, weigh, grade, process, and store the grain before shipping it for export or domestic consumption as per the buyer's specifications.

Grain quality is evaluated by several parameters such as physical factors (moisture content, bulk density, kernel size, kernel hardness, vitreousness, kernel density, and damaged kernels), sanitary factors (fungal infection, mycotoxins, insects and mites and their fragments, foreign material, odor, and dust), and intrinsic factors (milling yield, oil content, protein content, starch content, and viability) (Muir, 2001). Canadian wheat is graded based on the bulk density, varietal purity, vitreousness, soundness, and foreign material content (CIGI, 1993). There is an increasing trend among grain buyers towards zero-tolerance to grain contamination. Canada has a zero tolerance policy in place for live

insects in grain (Canada Grains Act, 1975) and a strict grain grading and inspection process. The Canadian grain industry is regulated by the Canadian Grain Commission (CGC, 2006). At terminal elevators CGC staff supervise and coordinate extensive grain monitoring, as the grain is received and shipped and support the grain certification procedure. At present, manual or chemical inspection methods are used for grain quality evaluation at the grain handling facilities, which are less efficient and time consuming than automated methods. Therefore, a rapid, objective, and accurate technique is needed for grain quality estimation in real-time which meets the standards of the international grain market. However, it is not possible to investigate all the aforementioned quality attributes in one PhD thesis. Therefore, in the present study insect infestation, fungal infection, and sprout damage in wheat were selected as grain quality parameters to be investigated. These infestations and damage result in quantity and quality losses, reduce the grain grade, and thus lower the market price of the grain. Detecting the presence of these infestations and damages is an important factor in the integrated approach to stored-grain quality control.

Canada loses a significant amount of grain due to stored-product insects and associated spoilage and remedial actions (White, 1995). Stored grain can have losses in both quantity and quality during storage as a result of insect or mite infestation, fungal infection or any combination of these. Deterioration and contamination result in downgrading of grain due to the presence of insect parts and odor, grain weight loss, increase in free fatty acids, heat damage, toxicity due to fungal activity, and poor milling and baking quality. Insect infestation in wheat adversely affects the baking quality of the flour (decreased loaf volume, compact and inelastic crumb, bitter taste, and off-flavors)

(Sanchez-Marinez et al., 1997). Insect damage to grains results in loss of nutrients, germination ability, and increased susceptibility to contamination by fungi. Insects produce heat and moisture due to their metabolic activity, which can cause development of insect-induced localized hotspots in grain bins and spoilage of grain by fungi. Detection of internal infestation by insects such as weevils and borers in the whole grain is complicated by the presence of hidden immature stages (eggs, larvae, and pupae) inside grain kernels. Sometimes samples of grain may appear to be insect-free due to the absence of adults; however, they might be infested by hidden immature stages.

The Canadian Grain Commission tests the grain samples for mould damage. Mouldy kernels are defined as discolored, swollen, and soft as a result of decomposition by fungi or bacteria (CGC, 2006). Mould or fungal growth in grains results in germination loss, discoloration, dry matter loss, increase in free fatty acids, heating, mustiness, and production of mycotoxins. Mycotoxins are toxic carcinogens which produce a serious health hazard such as acute or chronic liver disease and liver cancer to humans and animals that consume the infected grain or grain products (Tseng et al., 1995). *Penicillium* spp. (blue) and *Aspergillus* spp. (black, white, brown, blue, green, and yellow) are the most common storage fungi found in grain. There are more than three hundred mycotoxins mostly produced by *Penicillium*, *Aspergillus*, and *Fusarium* species.

Sprouting in western Canadian wheat is considered one of the important degrading factors and all wheat classes are tested for sprout damage. Sprouting is generally caused by germination of wheat kernels by absorbing moisture. Sprouted kernels can be easily contaminated with pathogens and are ideal for the growth of bacteria (Health Canada, 2006). These pathogens can be trapped in cracks and crevices

developed in the sprouted kernels and can cause cross-contamination when the grain is being mixed in grain elevators, shipped, or at processing facilities. Stored grain with a significant level of sprout damage is more vulnerable to infestation by insects as they can easily feed on the damaged kernels. Pre-harvest sprouting occurs mostly due to rain water when the wheat is left in the swath after cutting. Wheat breeders have developed new lines which are very resistant to pre-harvest sprouting. However, wheat midge infestation is still a serious cause of sprouting, and yield and economic losses. Oakley (1994) first hypothesized that midge damage splits the kernel's pericarp, facilitating the water uptake and hence sprouting in poor weather. Then Lunn et al. (1995) also concluded that sprouting of grain is the result of interaction between midge-damaged kernels and weather conditions. Midge damage is caused in the early development stage of wheat seeds in the field due to attack of the orange wheat blossom midge (*Sitodiplosis mosellana* Gehin) larva which feeds on the soft kernels (Oakley et al., 1998) and kernels become shrivel, cracked, deformed and misshapen. Alpha-amylase, an enzyme, which is present in very high concentration in sprouted wheat kernels (Kruger, 1994) affects the baking quality of bread, and thus lowers the premium paid for wheat. Enzymatic activity of alpha-amylase in the sprouted wheat kernels results in decrease in starch, increase in sugar, increase in total protein, change in amino acid composition, and dry matter loss (Lorenz and Valvano, 1981). The grain buyers, who use wheat to make flour for bread production, look for several parameters such as water absorption capacity, dough mixing, loaf volume, crumb strength, and slicing characteristics of baked bread. These quality parameters are adversely affected due to starch damage in sprouted wheat kernels (Fenney et al., 1988).

Near-infrared (NIR) spectroscopy has evolved as a fast, reliable, accurate, and economical technique available for grain quality analysis (Kim et al., 2003). The NIR region of the electromagnetic spectrum has absorption bands associated with wheat protein, other kernel compositions, and functionality (Pasikatan and Dowell, 2004). This multi-analytical technique is rapid, requires very little or no sample preparation, and several parameters can be estimated simultaneously. Near-infrared spectroscopy has been used for quality evaluation of many cereal grains (Singh et al., 2006). However, conventional NIR spectroscopic instruments are considered as point-based scanning instruments as they provide only one spectrum of the target sample without having any information about the distribution of the chemical composition of a sample. The compositional information is important in the grain due to its non-uniform distribution among the grain kernels. The NIR spectroscopic detection methods also have certain drawbacks such as the imprecise nature of the estimates, complex development of robust calibration models, and inconsistency across several individual instruments (Toews et al., 2007). Most of the commercially available NIR instruments are too insensitive for practical use in the field (Davies et al., 2003). Since the commercial NIR spectroscopy instruments are used for point measurements instead of whole kernel surface, multispectral or hyperspectral imaging systems are required (Nicolai et al., 2006).

Color images have been used in grain quality analysis to identify different grain types, varieties, classes, impurities, fungal-infected, and insect-damaged kernels (Zhang et al., 1999; Paliwal et al., 2003; Choudhary et al., 2008). Color images of individual kernels are described by color, textural, and morphological features and are used in quality assessment of grain. Color cameras are the least expensive and widely used in

machine vision systems in many industrial inspection applications. However, color images do not provide information about the chemical composition and its distribution in the kernel and rely mainly on external surface features of the object under investigation.

Hyperspectral imaging provides the spectral information in a spatially resolved manner. Spatial information is important for monitoring of the grain as it can be used to extract the chemical mapping of the sample from a hypercube. The mappings can be visually examined by grain inspectors without any subjectivity. The hyperspectral imaging has been used for a long time in remote sensing for earth resource monitoring and is finding new applications in medicine, agriculture, manufacturing, and forensic (Gat, 2000). The application of hyperspectral imaging has been evaluated for quality inspection of several agriculture products (Lu, 2003). In hyperspectral imaging, the hyperspectral data can be represented quantitatively in three forms, i.e., image form, spectra form (absorbance or reflectance), and feature form (discrete points in N-dimensional vector space), however, feature form is most commonly used to develop supervised classification models.

In NIR hyperspectral imaging, sensors working in the long-wave NIR (LWNIR) region (900-1700 nm) produce very high quality data with very high signal-to-noise ratio. However, the cost of these sensors (e.g., InGaAs) is very high. Hyperspectral imaging systems working in the short-wave NIR (SWNIR) region (700-1100 nm) (e.g., CCD) are relatively less expensive and color imaging systems are the least expensive among these three systems. The goal of the development of any quality monitoring system would be to achieve a higher level of performance at a lower system cost without compromising the system performance. The overall objective of this research was to assess the potential of

LWNIR hyperspectral imaging for detection of insect infestation, fungal infection, and incidence of sprouting in wheat, and compare the performance with SWNIR and color imaging. The main objectives of this research were to:

1. develop algorithms to detect insect and fungal damage, and incidence of sprouting in wheat using LWNIR hyperspectral imaging;
2. develop algorithms to detect insect and fungal damage, and incidence of sprouting in wheat using SWNIR hyperspectral imaging;
3. develop algorithms to detect insect and fungal damage, and incidence of sprouting in wheat using color digital imaging;
4. combine the features from color and SWNIR imaging and develop classification algorithms; and
5. Compare the classification results of hyperspectral imaging with color imaging.

## 2. THEORETICAL BACKGROUND

### 2.1 Electromagnetic Spectrum

The electromagnetic spectrum represents the distribution of electromagnetic radiation (continuum of energy packets called photons) in a one-dimensional scale spanning from very short wavelengths of gamma radiation (very high frequency and energy) to the long wavelengths of radio waves (low frequency and energy) (Fig. 2.1). Electromagnetic radiation is an energy wave (photon energy) composed of both electric and magnetic field properties which interact with matter to form a spectrum. Near-infrared region is the wavelength region from 700 to 2500 nm (Wavenumber 14300 to 4000  $\text{cm}^{-1}$ ) between the visible (VIS) and the infrared (IR) radiation (Osborne et al., 1993) (Fig. 2.1). William Herschel, an astronomer, discovered the IR spectral radiation in 1800 while investigating the cause of the heating of his telescope. The visible region covers from 400 to 700 nm of the electromagnetic spectrum and is the only region of electromagnetic radiation to which human eyes can respond. Electromagnetic radiations in gamma and X-ray waves have very high energy which may damage biological samples, so these wavelengths have limited applications for the analysis of biomaterials. The ultraviolet (UV), VIS, and IR radiation interact mainly with the surface of a material and have little penetration power. Ultraviolet radiation is useful for detection of fluorescence in agriculture products. Visible radiation is widely used in color-based grain sorting machines. Due to low penetration power of IR radiation, it is mainly used for detection of molecules in gases (Lestander, 2003). Microwave and radio waves have high transmission power to pass through biomaterials and have been used in grain disinfestations (Vadivambal et al., 2007) and moisture distribution study in grains using magnetic resonance imaging (MRI) (Ghosh et al., 2007). Near-infrared radiation has

higher penetration power than UV, VIS, and IR radiation and can be used in non-destructive analysis of both solids and liquids.

Spectroscopy is the study of physical characteristics of atoms or molecules using electromagnetic radiation (spectrum) in the form of absorption, emission, or scattering by molecules. Electromagnetic radiation has both wave and particle characteristics. When a beam of light is directed onto the surface of a material, a part of the incident light is

**Fig. 2.1 Electromagnetic spectrum (Lestander, 2003)**

absorbed by atoms or molecules exciting them to higher energy levels depending on the wavelength and intensity of the light source. This process is called absorption. The remainder of the incident light is either reflected or transmitted through the material. The excited atoms or molecules can return to lower energy levels by radiating the absorbed energy (emission). These different energy levels can be distinguished by integers called quantum numbers. However, a transition in the energy levels can only occur if the appropriate amount of energy (discrete packets of energy called quanta) is either being absorbed or emitted by the atoms or molecules. A molecule in space can have several

forms of energy such as vibrational, rotational, electronic, or translational. Vibrational energy is the result of periodic displacement of atoms from their equilibrium position and rotational energy is the result of rotation about a center of gravity. The energy required for rotational state is low (far infrared and microwave region) compared to the energy required for change in vibrational state (NIR-mid IR). In general, rotational spectra exist in gases and only vibrational motion is considered in the spectral analysis of solids and liquids (Osborne et al., 1993). The following section discusses the basics of the mechanical model and different forms of energy.

### 2.1.1 Mechanical models

**2.1.1.1 Harmonic model** The simplest mechanical model of diatomic molecular vibration consists of two spherical masses ( $m_1$  and  $m_2$ ) connected with a spring of given force constant ( $k$ ). According to the Hook's law, the energy ( $E$ ) of this system is given by:

$$E = \frac{h}{2\pi} \sqrt{\frac{k}{\mu}} \quad (2.1)$$

where,  $h$  is Planck's constant and  $\mu$  is the reduced mass given by the following equation:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (2.2)$$

A system having more than two masses will follow the same principle. The molecular vibration can be considered as having a harmonic oscillatory path with nearly zero potential energy at the equilibrium (Fig. 2.2A) (Pasquini, 2003). The potential energy

( $U$ ), as a function of the displacement of the atoms ( $x$ ), is expressed by the following relationship:

$$U = \frac{1}{2} kx^2 \quad (2.3)$$

This mechanical model is useful in understanding the concept of vibrational energy and in explaining many spectral observations. A covalent bond between two atoms in a molecule follows the same principle. If a covalent bond is stretched from the equilibrium position, the electromagnetic force between the electrons in the covalent bond becomes weaker and breaks when the potential energy reaches the bond energy. Harmonic model is suitable for smaller displacements from equilibrium. In such cases the vibrational displacement is governed by attractive and repulsive electromagnetic forces.

**2.1.1.2 Quantum mechanical theory** Planck proposed that the molecular systems cannot assume the continuous energy states and any change in the oscillation energy can take place by transition between two discrete energy levels. According to Planck's theory this transition energy change can take the form of electromagnetic radiation and the frequency of radiation is related to energy change. The discrete energy levels ( $E_\nu$ ) of the molecules can be expressed by the following equation:

$$E_\nu = \left(\nu + \frac{1}{2}\right)h\nu \quad (2.4)$$

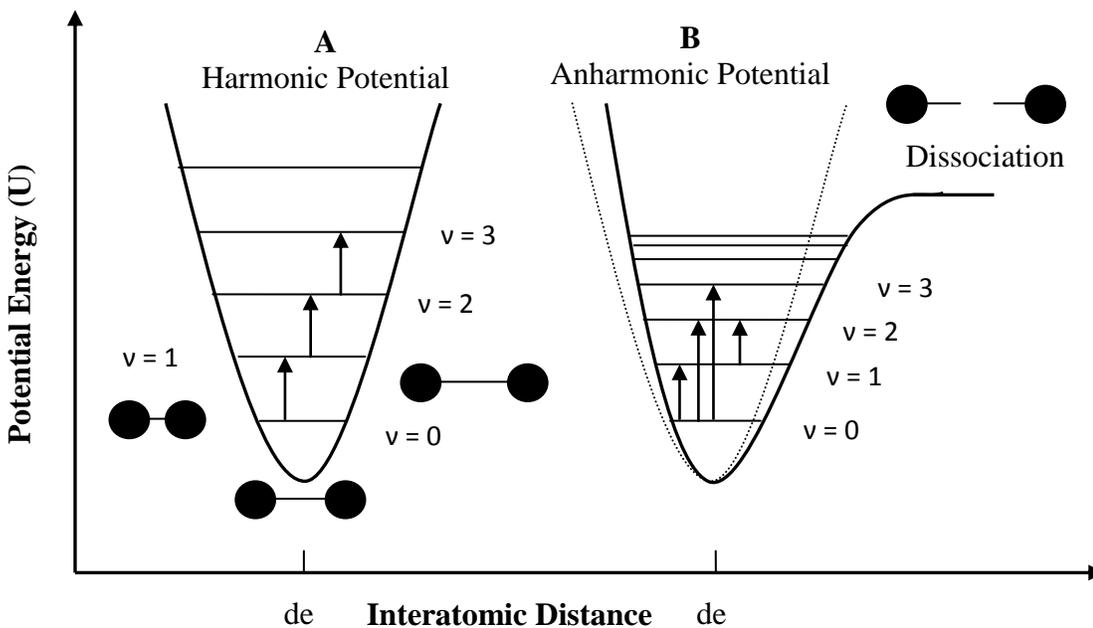
where,  $\nu$  is the quantum number,  $E_\nu$  is the energy associated with the  $\nu^{\text{th}}$  quantum level, and  $\nu$  is the vibrational frequency. In the harmonic model this fundamental frequency was defined in equation (1) as:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \quad (2.5)$$

Transition of the energy in the quantum mechanical model can only occur between adjacent energy states,  $\Delta v = \pm 1$ . If the photon energy of the electromagnetic radiation is equal to the bond energy of a diatomic molecule then the photon is absorbed and the molecule jumps from the ground state to an excited state instantly causing vibration. The molecule releases the absorbed energy as heat and returns to the ground state. The energy level required to excite a molecule from a lower to a higher energy level is thus given by:

$$\Delta E = E_{v_2} - E_{v_1} = \Delta v h \nu \quad (2.6)$$

where,  $\Delta E$  is energy difference.



**Fig. 2.2 Diatomic harmonic (A) and anharmonic (B) models and their potential energy (Pasquini, 2003)**

The photons in the IR region have lower energies that correspond to covalent bond stretch and bend vibrations which can cause only fundamental transition from ground

state to first excited state (single band). An absorption spectrum can be produced by collecting the radiation after its interaction with the material. Fundamental vibrations are studied in spectral IR analysis. In the NIR region more than one absorption band are found which cannot be explained by the harmonic model due to the assumption of equally spaced discrete energy levels.

**2.1.1.3 Anharmonic model** The quantum model is able to explain the vibrations occurring in the IR region due to fundamental modes but can not explain the overtones of the NIR region. An anharmonic model can explain this, which is based on the phenomenon that two nuclei cause a rapid increase in potential energy compared to a harmonic model specifically when these are stretched. This asymmetric distribution results in quantized but unequally spaced energy levels, therefore, absorption bands are not the multiple of fundamental absorption bands. Morse function describes the anharmonic behavior of a diatomic molecule (Fig. 2B) (Pasquini, 2003) using the following equation:

$$U = E_d(1 - e^{-\alpha y}) \quad (2.7)$$

where  $\alpha$  is a constant for a specific molecule,  $E_d$  is the dissociation energy, and  $y$  is the displacement.

Quantum mechanical model for anharmonic oscillation can be derived using Morse equation as:

$$E = ( (\nu + 0.5)h\nu - (\nu + 0.5)^2 h\nu x_0 - (\nu + 0.5)^3 h\nu x_1 - \dots ) \quad (2.8)$$

where  $x_0, x_1 \dots$  are the anharmonicity constants of the vibration, whose values are between 0.005 and 0.05 in descending order. For smaller values of  $\nu$  the third term and onwards can be ignored resulting in:

$$E = h\nu ((\nu + 0.5) - (\nu + 0.5)^2 x_0) \quad (2.9)$$

For fundamental vibration, i.e. from  $\nu=0$  to  $\nu=1$ , the net transitional energy can be derived from the above equation, which is given by

$$E = h\nu (1 - 2x_0) \quad (2.10)$$

The anharmonic model can also predict the transition energy for transition from  $\nu=0$  to higher transition levels of  $\nu=2$  and  $\nu=3$ , which correspond to first and second overtones, respectively. In the anharmonic model, the vibrations of the molecules depend on each other and can interact with them. Combination bands occur as a result of combination of overtones and fundamental vibration in various combinations. Overtones and combination bands are weaker than fundamental absorptions. This is an advantage with NIR radiation over IR radiation and samples with several mm or cm thicknesses can be analyzed.

## 2.2 NIR Spectroscopy

Near-infrared spectroscopic instruments record the absorption of NIR radiation by a material at a specified wavelength range with a narrow resolution. In NIR instruments, signal to noise ratio is very high (10000:1) (Hans, 2003). The NIR spectroscopic technique works on the principle that unique chemical composition of the matter causes molecules to absorb light in the NIR region and vibrate at unique frequencies (Murray and Williams, 1990). Reflected or transmitted light is collected by a spectrometer (detector) and related to the energy absorbed by the sample by transforming the reflected or transmitted radiation into absorbance. The NIR spectrum shape is characterized by overtones and combination bands of fundamental vibrations occurring in the NIR region. Near-infrared spectra obtained from the absorption or reflection/transmission at multiple

wavelengths can be related to the concentration of a particular constituent of the sample. Due to complex molecular structures of most organic compounds, the NIR spectra arising from overtones and combination bands have broad and highly overlapping peaks and valleys. This makes spectral analysis and extraction of chemical and physical information from the spectra very difficult. Chemometric methods have been widely used as qualitative and quantitative analytical tools in NIR spectroscopy. In spectroscopy, single-point absorption is recorded per sample by detector at each wavelength.

### **2.3 Color Imaging**

Color, which is produced by single wavelength or a narrow band of wavelengths in the visible region (400-700 nm) is called monochrome color. The image developed by a two-dimensional detector by sensing the reflection of the light of single wavelength is called a monochrome image. A typical monochrome image is described by two-dimensional array of size  $m \times n$ , where  $m$  and  $n$  are the pixels (detector size) in the  $x$  and  $y$  dimensions, respectively. A color image is created by combining the monochrome images corresponding to red (R), green (G), and blue (B) wavelengths into a three-dimensional multispectral image. This three-dimensional multispectral RGB image can be described by an array of  $m \times n \times 3$  size, where  $m$  and  $n$  are detector size (pixels) in  $x$  and  $y$  spatial direction and the third dimension is spectral or wavelength direction comprising of R, G, and B intensities. The color information in the RGB image is used for discrimination of various objects by human eyes based on similarities and contrast among the objects. The discriminative information from the images is more precisely extracted by pattern recognition by processing images, applying mathematical models to extract image features, and developing supervised classification algorithms.

## 2.4 Hyperspectral Imaging

In NIR spectroscopy, single-point absorption is recorded at contiguous wavelengths in 5 to 10 nm intervals between 400-2500 nm. The multispectral images provide spatial distribution ( $m \times n$ ) but are restricted to a few wavelengths, e.g., only three wavelengths channels, R, G, and B in RGB color images. In hyperspectral imaging, spatial data can be recorded at hundreds of contiguous wavelengths in the NIR region. The hyperspectral images can be described by three-dimensional array of size  $m \times n \times \lambda$ , where  $m$  and  $n$  are the spatial dimensions (detector size or pixels) in  $x$  and  $y$  direction and  $\lambda$  is the wavelength or third dimension. The three-dimensional hyperspectral data are also referred to as a hypercube. One of the critical issues of hyperspectral imaging is the large size of data (hypercube) and overlap of information in adjacent image slices. This makes the extraction of relevant information with greater discriminating potential from hyperspectral data cumbersome and challenging. Analytical tools in chemometrics combined with spatial image processing are used to reduce the dimensionality of hyperspectral data, select the significant wavelengths, extract the features, and then develop calibration and classification models.

## 2.5 Instrumentation

**2.5.1 Color imaging system** An area-scan color imaging system consists of mainly lighting source, sensor (analog or digital camera), digital converter for analog camera, frame grabber, and a data storage device (computer).

**2.5.1.1 Lighting source** The lighting system should be able to produce uniform, stable radiation of desired spectral range. Light emitting diode, halogen lamps, incandescent

lamps, and fluorescent lamps are commonly used as radiation sources in the VIS region of the electromagnetic spectrum.

**2.5.1.2 Sensor** The sensors are semiconductor devices which have stationary electrons in valence band and empty conduction band. The absorption of photon energy by a semiconductor excites the stationary electrons in valence band to conduction band where the electron moves freely. This movement causes a change in conductance and produces electric current proportional to the photon energy absorbed. The sensors are structured into one-or two-dimensional array and accumulated charge is converted into output signal. The most common technology is called the charge-coupled device (CCD) which uses an analog shift register to transfer the accumulated charge. There are four analog camera standard formats namely RS-170, National Television System Committee (NTSC), Consultative Committee on International Radio (CCIR), and Phase Alternating Line (PAL). The accumulated charge is measured as voltage or current and converted by a digital converter called a digitizer to digital signal that can be stored and processed in a computer. Images can be captured, processed, and analyzed with the help of a frame grabber (imaging board) attached to a computer (Cheng et al., 2003). In digital CCD cameras, the digitizer is built as a part of the camera circuitry. Complementary metal oxide semiconductor (CMOS) is another technique in which the sensors are structured in a way that each sensor in an array has its own readout circuit with no charge accumulation. The sensors sensitive to the VIS radiation in the range of 400 nm to 700 nm are used in color cameras, which produce an electric output signal corresponding to electromagnetic energy sensed in the form of light reflected from the surface of an object. Single-chip CCD sensors are used to produce the monochrome images of R, G, and B

band by changing the intensity of light using a filter (bandpass or filter wheel). Three-chip CCD sensors are also used in a color camera which receives R, G, and B intensity light simultaneously by beam splitting the light using dichroic coating on the surface of a prism.

**2.5.1.3 Communication interface** Four standard communication interfaces namely Parallel, FireWire (IEEE 1394), CAMERA Link, and GiGE VISION are used to transfer digital image data between camera and computer. Parallel cameras have a high data transfer rate but they require customized cables due to the lack of an interface standard. FireWire is a standardized interface but has lower data transfer speed. CAMERA Link uses a standard channel link chip which is used in camera and frame grabber for data transmission. Advanced personal computer bus systems such as peripheral component interconnect (PCI) express can handle fast data streams transferred by Camera Link cable. The GiGE VISION is the latest developed standard interface with very high data transfer rate but can be used in limited bandwidth. FireWire and GiGE VISION communication interfaces do not require a frame grabber board and are directly connected to the computer.

**2.5.2 Hyperspectral imaging system** A hyperspectral imaging system mainly consists of a detector, wavelength selection device, radiation source (light), image acquisition software, and an integrated computer for data acquisition and storage. There are three types of hyperspectral imaging systems based on sample presentation techniques: point scan, line-scan (pushbroom), and area-scan. The selection of the above mentioned hardware components depend on the choice of imaging system and related application. The line- scan and area-scan imaging systems are better suited for food quality inspection

(Kim et al., 2001). In line-scan imaging, full spectral information for each pixel in one spatial dimension (line) is collected and successive line-scans are combined to form a three-dimensional hypercube. This system is suitable for scanning of moving objects. The scanning of moving object causes the geometric distortion in the images such as image registration which needs to be corrected. In area-scan imaging system, the spatial information is collected at each wavelength sequentially to form a three-dimensional array (hypercube). The first two dimensions of the hypercube represent spatial features (pixels) and the third dimension represents the spectral features (wavelength). This type of imaging system is used mainly to acquire the images of stationary objects.

**2.5.2.1 Detectors** Detectors record the sample spectra by selecting the suitable mode of energy wave (reflectance or transmittance). The single channel NIR detector (point scan) uses lead sulphide (PbS) in the range of 1100 to 2500 nm, silicon detectors in the range of 360 to 1050 nm, and Indium-Gallium-Arsenide detectors in the range of 900 to 1700 nm wavelengths. In hyperspectral imaging, the array of detectors which is also known as focal plane arrays (FPA) is commonly used instead of point based scanning to reduce scanning time, obtain uniform background, higher signal to noise ratios, and to overcome image distortion problems. In line-scan imaging, a linear array of detectors is used. In area-scan imaging, two-dimensional array of detectors is used. Silicon based CCD detectors are also available for SWNIR (700-1100 nm) imaging which are cost effective, however, they cannot be used for longer wavelength scanning application as silicon becomes transparent and need a coating material that reduces the quantum efficiency of the detector.

Different types of commercial FPAs currently available are: Indium Antimonide (InSb), Platinum Silicide (PtSi), Indium Gallium Arsenide (InGaAs), Germanium (Ge), Mercury Cadmium Telluride (HgCdTe), and quantum well infrared photodetectors (QWIPs). Out of these FPAs, InSb, InGaAs, HgCdTe, and QWIP are the most commonly used detectors in multispectral/hyperpectral imaging (Tran, 2003).

Indium antimonide (InSb) detectors have high quantum efficiency, wide spectral response from 1000-5000 nm, and lower band gap energy (0.22 eV at 77 °K). Due to lower band gap energy InSb detectors are highly sensitive and give very high quality uniform images but require cryogenic cooling to operate under room temperature. This makes the InSb FPAs very expensive. The InSb detectors can be a good choice in mid-IR based imaging systems such as thermal camera due to their low band gap energy, high sensitivity, and large spectral response in the mid-IR region.

Mercury Cadmium Telluride (HgCdTe) detectors also have high sensitivity and detectors can be designed to operate in very wide range of the IR region (2,000-26,000 nm) and cut-off can be selected by changing the composition of the material. These detectors can operate at room temperature. The biggest drawback with HgCdTe detectors is the instability and non-uniformity of pixels caused by high Hg vapor pressure during the material growth and thermal expansion mismatch (Tran, 2003).

The quantum well infrared photodetectors (QWIPs) are also used in long wave IR imaging systems. Their performance is better than HgCdTe at low temperature (45 °K) but produce high dark current as temperatures increase. Their low quantum efficiency is another disadvantage.

The InGaAs detectors are most commonly used in the 900-1700 nm wavelength band and have a large range of applicability in agriculture and food material inspection. The InGaAs detectors have high quantum efficiency in the 900-1700 nm wavelength interval (>85%). They have higher band gap energy compared to the InSb detectors which produce very low dark current and can be operated at room temperature by thermoelectric cooling. The InGaAs detectors have better sensitivity, wider spectral range, low noise, and faster response in the NIR region. The disadvantage of the InGaAs detectors is their poor sensitivity beyond 1700 nm range.

**2.5.2.2 Wavelength filtering devices** Wavelength filtering devices such as optical interference filters, grating devices (e.g., prism-grating-prism), and electronically tunable filters (ETF) are used to obtain the light of the desired wavebands and remove out-of-band radiation. The use of a filtering device is also highly dependent on the type of hyperspectral imaging system. In pushbroom type systems (moving objects on belt conveyer or moving camera in air-borne systems) grating devices are commonly used. In area-scan imaging (stationary objects) ETF are considered the most suitable filtering devices for sequential wavelength scanning. A filtering device should have the following characteristics such as: minimal tunability time, minimal out-of-band transmission, minimal physical thickness, low power consumption, insensitive to polarization, selectable bandpass, insensitive to environment (e.g., ambient temperature and day light fluctuations), insensitive to angle of incidence of the incoming light (wide field of view), infinite spectral range, large aperture, constant bandpass, random access to wavelengths (Gat, 2000). However, finding all these attributes in one type of ETF is not feasible and one should look for the optimized parameters for intended application while selecting the

filter. Acousto-optical tunable filter (AOTF) and liquid crystal tunable filter (LCTF) are two advanced ETFs and both have relatively large optical aperture, high spectral resolution, wide spectral range, and randomly accessible tuning wavelengths (Wang and Paliwal, 2007). These filters also do not have any moving parts which overcome the registration problem (distortion in image) in acquired images and is another advantage over interference filters and grating devices.

The AOTFs are acousto-optical electronically tunable devices based on diffraction in which wavelengths are selected by applying radio frequencies (RF) acoustic waves to a crystal material (Quartz,  $\text{TeO}_2$ ,  $\text{Ti}_3\text{AsSe}_3$ ) and wavelength of light produced is proportional to the RF frequency applied. The light of different wavelengths is selected by changing the RF frequency applied to the crystal and for a given frequency only the light corresponding to that frequency is diffracted from the crystal. Bandwidths as narrow as 1 nm full width at half maximum (FWHM) can be produced by AOTFS, however, the field of view (FOV) through AOTFs is relatively smaller than LCTFS (Call and Lodder, 2002). The AOTFs give the fastest scanning speed as they can be tuned in just a few microseconds by changing the frequency of acoustic waves. The AOTFs have compact size, all-solid-state design, rapid scanning speed, wide spectral tuning range (VIS-IR), high throughput (>90% diffraction efficiency), high spectral resolution, random and sequential wavelength filtering, and relatively lower power consumption (Tran, 2003).

The LCTFs are birefringence-based electronically tunable devices which are built using a stack of polarizers and tunable retardation (birefringent) liquid crystal plates (Tran, 2003). The birefringent liquid crystal is placed between two polarizers whose axes are parallel to each other. The incoming unpolarized light is converted into linearly

polarized light by input polarizer. The polarized light is then passed through the birefringent crystal and split into ordinary and extraordinary beams and a phase delay between the beams is introduced by the birefringent retarder (quartz or calcite). The optical path difference between these two beams is called retardance. The light coming out of the retarder then passes through a second polarizer which blocks out-of-band transmission. A LCTF is constructed and electrically tuned by placing the birefringent retarders in series with the liquid crystal variable waveplates which are in order of increasing retardance. The LCTFs are being widely used in the area-scan hyperspectral imaging systems.

**2.5.2.3 Illumination sources** Selection of appropriate illumination (lighting) source is one of the most important factors in producing high quality and noise-free images. The illumination source should be able to produce the light sensitive to the camera in the desired wavelength application range. Tungsten halogen lamps, quartz halogen lamps, light emitting diodes (LED), and tunable lasers are used as light sources in NIR instruments. Heated xenon lamps can also be used as sources of illumination in NIR instruments. The application of LED is restricted to only narrow-bands (400-900 nm) of wavelengths. Tungsten halogen lamps are the most common illumination source in NIR hyperspectral imaging system due to their durability, stability, and capability to emit light in broad spectral range of 400-2500 nm.

**2.5.2.4 Integration of hardware and software** The image data captured by the NIR detector are transferred to a computer for storage and analysis with the help of earlier described standard communication interfaces, e.g., FireWire, CAMERA Link, and GiGE

VISION. Various calibration, pre-processing, data reduction, and classification methods are used to analyze hyperspectral data.

## **2.6 Data Analysis**

A number of chemometric tools (mathematical and statistical techniques) are available and used for extracting relevant information from NIR spectroscopic data to develop classification models. The large spectral data with presence of highly overlapping, broad, and low absorption bands need to apply an appropriate data reduction method. Principal component analysis (PCA) is the most common data reduction technique used in spectroscopic data analysis. In a recent review paper Van der Maaten (2007) compared the performance of two linear techniques, linear discriminant analysis (LDA) and PCA with non-linear techniques namely multidimensional scaling (MDS), isomap, kernel PCA, diffusion maps, multilayer autoencoders, local linear embedding, laplacian eigen maps, Hessian local linear embedding, local tangent space analysis and local linear coordination on natural and artificial data. The authors in this comparative review concluded that non-linear techniques were not able to outperform the linear dimension reduction techniques. The linear techniques are simple and carry less computation burden. Wang and Paliwal (2006) applied PCA, Fourier transform (FT), and wavelet transform (WT) to reduce the dimensionality of NIR spectral data and developed classification algorithms using LDA, quadratic discriminant analysis (QDA), k-nearest neighbour (KNN), probabilistic neural network (PNN), and least-squares support vector machines (LS-SVM). They found PCA was the most efficient in compressing the spectral data and LDA using PCA scores gave the highest classification accuracy. These

techniques have also been used in two-dimensional image data compression and feature selection.

Detailed calibration procedures have been developed for line-scan and area-scan hyperspectral imaging systems by Lawrence et al. (2003a) and Geladi et al. (2004), respectively. Multivariate image analysis (MVI) techniques applying PCA are used for data reduction and feature extractions from hyperspectral images. The PCA technique maximizes the image features with optimum number of independent components by orthogonally transforming the original hyperspectral data. Wavelet transform has also been applied in hyperspectral imaging. Independent component analysis (ICA), also known as blind source separation technique, has also been applied in hyperspectral imaging, however, the disadvantage of this method is that it requires the number of independent components to be given or specified before the computational analysis. In PCA the number of significant components can be determined from the variance, score, and factor loadings observation after the computation is completed. A number of methods applied to hyperspectral image data for dimension reduction, wavelength selection, classification, and prediction include band ratio, band difference, partial least square (PLS), PCA, PCA- Fisher's discriminant analysis (FDA), multi linear regression (MLR), artificial neural network (ANN), decision tree, spectral angle mapper, hybrid PC-ANN, fuzzy logic, support vector machines (SVM), image thresholding (Gowen et al. 2007).

### 3. SYNTHESIS OF LITERATURE

#### 3.1 Detection of Insect, Fungal, and Sprout Damage

**3.1.1 Insect damage** Presently, the Berlese funnel method is used by the Canadian Grain Commission as a standard method to detect live insects in grain samples. Heat from a light source drives insects out of 1kg samples of a grain into a collecting jar. Traditional insect detection methods such as visual inspection, sieving, insect traps, insect fragment inspection, and floatation methods are commonly used for the detection of insect contaminants in grain. Most of these methods have one or more drawbacks such as being subjective, destructive, inaccurate, time consuming, and sometimes unable to detect insect infestation in side kernels. Methods such as enzyme-linked immunosorbent assays (Brader et al., 2002), carbon dioxide and uric acid measurement (Karunakaran et al., 2004), electronic nose (Zhang and Wang, 2007), acid hydrolysis (Brader et al., 2002), electrical conductance (Pearson and Brabec, 2007), MRI (Chambers et al., 1984), computed tomography (Toews et al., 2006), acoustic impact emissions (Pearson et al., 2007), machine vision (Zayas and Flinn, 1998), and thermal imaging (Manickavasagan et al., 2008) have also been investigated for insect detection in grains. Most of these methods, however, are unable to detect the low-level internal infestations and have not shown the potential for automated inspection. For detection of insect-damaged grain at grain handling facilities, an objective, non-destructive, rapid, and accurate method is required.

Near-infrared spectroscopy has been used for detection of insect and insect parts in whole grain and ground samples (Dowell et al., 1998; Baker et al., 1999; Dowell et al., 1999; Maghirang et al., 2003). Ridgway and Chambers (1998) used hyperspectral

imaging for detection of insect damage in wheat but did not develop any supervised classification algorithm for future prediction. In another study, Ridgway et al. (1999) used the spectroscopic data to select the most important wavelengths and then used the image differences (at few selected wavelengths) for detection of insect damage. Hyperspectral images can be directly analyzed for wavelength selection using multivariate image analysis without using spectroscopic data to overcome system variations between these two instruments.

**3.1.2 Fungal damage** In Canada, there is a single tolerance for the total of binburnt, severely mildewed, mouldy, and rotted kernels. The grain samples are inspected for mould damage by grain inspectors, which is subjective and time consuming. Some species of insects and mites feed on fungi and make it more difficult to detect the fungal infection. Current methods being practiced for fungal detection are: chemical (Sukurai and Shiota, 1977; Seitz et al., 1977, 1979; Gaunt et al., 1985; Sashidhar et al., 1988), immunological (Notermans et al., 1986; Kamphuis et al., 1989; Notermans and Kamphuis, 1992), selective and differential media (Gourama and Bullerman, 1995), fungal volatiles (Borjesson et al., 1993; Zeringue et al., 1993; Scotter et al., 2005), most probable number (Brodsky et al., 1982), Howard mold count, and polymerase chain reaction (Manonmani et al., 2005). In chemical methods, proteins and carbohydrates of the cereals may interfere with the fungi which may lead to misinterpretation of the results. Microbial cultural methods which are traditionally used for fungal detection require long incubation periods (Magan, 1993). Microscopic detection and quantification of fungi is time consuming and not suitable for rapid applications (Williams, 1989). Aflatoxin-producing fungi are also qualitatively identified using aflatoxin fluorescence

response (Zeringue et al., 1999; Malone et al., 2000) which is time consuming and inaccurate (Yao et al., 2005).

Research on non-invasive techniques for fungal detection has also been conducted. Various techniques such as ultrasound (Walcott and McGee, 1998), machine vision (Paulsen, 1990), electrical impedance (Jarvis et al., 1983) have been investigated for their application in fungal detection in grain. Gordon et al. (1997) used Fourier transform photoacoustic infrared spectroscopy to detect fungal damage in corn kernels and correctly classified all the healthy and infected corn kernels. However, they visually selected the top ten spectral features and manually assigned a class to the kernels based on spectral feature differences using a very small sample set (20 kernels). For practical use, a more versatile sample, statistical computation, and supervised model development for future prediction is required and should be validated. Color imaging was also used to develop classification models to detect fungal-damaged soybean seeds which gave 77-91% classification accuracy (Casady et al., 1992, Ahmad et al., 1999). Ridgway et al. (2002) used a monochrome CCD camera to detect a toxic fungal body known as ergot in bulk wheat and rye samples and achieved a classification rate of 87%. Wang et al. (2003) used NIR spectroscopy to detect fungal damage in soybean seeds and developed two-class PLS and five-class ANN models. Using visible (490-750 nm) or NIR (750-1690) region alone gave lower classification, whereas, combining visible and NIR region (490-1690) gave higher accuracy and correct classification of 84-100%. Paolesse et al. (2006) studied the application of an electronic nose to detect the fungal infection in the cereal grains. They artificially infected the grains and used small sample size (50 g) in a bottle to measure the volatile composition in the head space. So their study did not replicate the

practical grain storage conditions where grain is stored in very large bins. The results from their chemometric analysis of data showed poor discrimination capability between the grain samples having the same water activity (grain moisture) as fungi overlapped between classes and reduced the classification accuracy. Complex chemical pattern of volatile components make the analysis difficult and time consuming. Naravankar et al. (2009) used x-ray imaging to develop supervised classification algorithms to detect fungal damage in wheat kernels and correctly classified 82% healthy and 92.2-98.9% fungal damaged kernels. However, their models failed to distinguish between the kernels infected by individual fungal species. Most of non-destructive methods are not able to detect the grains with a relatively low level of infection where clear physiological changes are not visible. Therefore, it is necessary to develop a method for the detection of fungal infection which will be fast, reliable, nondestructive, user friendly, and easily upgradable.

**3.1.3 Sprout damage** All western Canadian wheat classes are visually examined for sprout damage by the Canadian Grain Commission inspectors. A traditional method in which falling number is related to alpha-amylase is used for evaluation of sprouted kernels in a bulk sample (Approved method 56-81B, AACC International, 2000). Rapid visco-analyzer is also used for the screening of sprouted kernels (Approved method 22-08, AACC International, 2000). The visual inspection methods are either subjective or inconsistent and slow, and chemical methods are destructive and time consuming (Neethirajan et al., 2007; Mares, 1993), so they can not be used for online inspection. Neethirajan et al. (2007) used X-rays to classify sprouted and healthy wheat kernels and identified 95% sprouted and 90% healthy kernels; however, X-rays pose potential health

risks to humans due to exposure to ionizing radiation from a defective system. Some of the millers and bakers demand wheat with a very low level of sprout damage. Therefore, an accurate and timely detection technique for sprout-damaged wheat kernels is needed. Near-infrared spectroscopy has been applied for quality evaluation of many cereal grains (Singh et al., 2006). Shashikumar et al. (1993) used a NIR reflectance analyzer to predict sprout damage in wheat. Instead of using the whole NIR wavelength range to develop the calibration model, they used only three different combinations of wavelength filters in the range of 1445-2345 nm. This lack of robustness in the calibration model did not give very good prediction accuracy ( $r = 0.75-0.87$ ) but demonstrated the potential for sprout damage detection using NIR spectra. Studies detecting artificially sprouted wheat kernels by hyperspectral imaging have also been reported (Koc et al., 2008; Smail et al., 2006). Most of these studies have used only PCA score images for detection of damaged kernels by unsupervised classification and did not develop any algorithm for future prediction by training the classifier.

### **3.2 Applications of NIR Spectroscopy**

Grain quality is defined by several factors such as moisture content, bulk density, kernel size, kernel hardness, vitreousness, kernel density, total damaged kernels, fungal infection, mycotoxins, insects and mites infestations, foreign material, and chemical composition (Singh et al., 2006). Most of these grain quality parameters can be determined using NIR spectroscopy. A thorough review on applications of near-infrared spectroscopy and imaging in food quality and safety was presented by Wang and Paliwal (2007). A considerable research reported on applications of NIR spectroscopy for grain

quality analysis is summarized in Table A.1 in the APENDIX. These applications are discussed in detail in this section.

Wheat is grouped into several classes based primarily on color, hardness, and growing season to get better market price as well as for various end use applications. Delwiche and Masie (1996) classified kernels of five wheat classes (hard white, hard red spring (HRS), hard red winter (HRW), soft red winter (SRW), and soft white) using visible and NIR reflectance characteristics obtained from single kernels. Their PLS based binary calibration models using both VIS and NIR regions gave accuracy of the 78-99%. Hard white and hard red winter wheat kernels were well differentiated in the VIS region (99% accuracy) as pigment color played a major role in classification. Accuracy of the five-class model ranged from 65%-92%. Grain color and appearance affect the market value of the grain and misclassification results in loss of monetary value if misclassification is to a lower grade, and potential exists for loss in reputation if the misclassification is to a higher grade for a buyer. Red wheat and white wheat have different milling, baking, taste properties, and different visual characteristics (Dowell, 1997).

Wang et al. (1999) classified wheat kernels based on kernel colors. They studied three different wavelength regions 500-750, 500-1700, and 750-1900 nm for wheat color classification. The spectral data were pretreated with first and second derivatives but they lowered the accuracy which could be due to an increase in noise. The PLS calibration model using absorbance spectra in the region 500-1700 nm correctly classified 99.8 and 98.4% kernels in a calibration set and a test set, respectively. The results showed that kernel color classification should not be restricted to the VIS region only. Ram et al.

(2002) classified wheat kernels using 490-750 nm wavelength region but they enhanced color contrast between red and white kernels before scanning the samples by bleaching them in NaOH solution (bleaching and soaking of the kernels delayed the analysis and removed the natural grain color). These problems could be possibly avoided by use of the VIS-NIR (400-1700) region for color classification.

Mohan et al. (2005) classified bulk grain samples of seven cereals namely, Canada Western Red Spring (CWRS) wheat, Canada Western Amber Durum (CWAD) wheat, Canada Western Soft White Spring (CWSWS) wheat, 2-row barley, 6-row barley, oats, and rye using VIS-NIR reflectance spectra. They extracted wavelength ratios and wavelength slopes from reflectance data and used these as input to linear parametric and back propagation neural network (BPNN) classifiers for classification of cereal grains. They correctly classified more than 99% of grain samples by developing calibration models using top five slope features.

Wheat grain hardness is a quality characteristic that defines or determines milling characteristics and its end use (Slaughter et al., 1992). However, HRS wheat is considered having higher baking quality as compared to HRW wheat. In standard hardness measurement methods (e.g., AACCI, NIST), the grain samples are ground before the analysis. Delwiche and Norris (1993) developed a calibration model to classify the hard red wheat into HRS and HRW wheat based on growing season. They applied Mahalanobis discriminant analysis to the scores of principal components and achieved maximum accuracy of 95%. Year-to-year changes in protein and hardness value affected the classification accuracy. Use of ground samples makes this analysis destructive. In another study, Delwiche et al. (1995) improved this model by analyzing bulk wheat

samples by scanning the same wheat samples used in previous study in the range of 1100-2498 nm. They evaluated the model performance by applying four classification algorithms, MLR, PCA-Mahalanobis, PLS, and ANN to second derivative diffuse reflectance spectra of bulk wheat samples. The ANN model gave the highest classification accuracy (up to 98%). The performance of this model was very high but it was not able to detect any other wheat class present in the bulk samples. Single kernel analysis of the samples could offset this problem by adding other wheat classes into model calibration. Maghirang and Dowell (2003) measured hardness of bulk wheat by calibrating the spectral data of single kernels in the range of 400-1700 nm. They analyzed the spectra by applying PLS on the averaged spectra in the 550-1690 and 950-1690 nm wavelength regions. More than 97% kernels were correctly classified as soft or hard wheat with 550-1690 nm wavelength range. Inclusion of visible range improved the performance of the model. Bulk samples containing more than 20% contrast hardness (e.g., 20% hard and 80% soft) gave 90-100% correct classification. A quantitative calibration model was also developed to predict the wheat hardness by measuring the reference hardness value by SKCS 4100 system (Perten Instruments, Springfield, IL), which predicted the hardness of soft and hard wheat with high accuracy ( $r^2=0.88$ ).

Vitreousness of hard wheat is the glossy or shiny appearance of the wheat kernel. Vitreousness affects the milling performance of durum wheat (semolina yield) and quality of some products from semolina (e.g., pasta). It is an indicator of hardness and high protein content and is considered by the wheat industry as one of the milling and cooking quality parameters. Current visual inspection method to determine vitreousness by grain inspectors is subjective and time consuming. Dowell (2000) used NIR

spectroscopy (400-1700 nm) to classify single wheat kernels into vitreous and non-vitreous classes. Classification algorithms were developed by PLS using mean centered diffuse reflectance spectra. All of obvious vitreous and nonvitreous kernels were correctly classified, however, classification accuracy reduced to 75% when difficult to classify kernels were included in the test set. He also investigated the effect of protein, starch, and hardness on vitreousness classification by excluding wavelength regions 750-1250 nm and 1400-1700 nm (significant for protein, starch, and hardness) from calibration and found that information outside these wavelength regions is useful for vitreousness classification. Wang et al. (2002) further sub-classified HRS wheat in dark hard vitreous (DHV) and non-dark hard vitreous (NDHV) kernels by analyzing the spectra from single DHV and NDHV kernels obtained using a diode-array spectrometer. The HRS wheat samples included checked, cracked, sprouted, and bleached kernels. Their PLS model gave good prediction results (97.1-100%) but failed to classify bleached kernels. They also investigated the effect of wavelength regions (500-750, 500-1700, and 750-1700 nm) and kernel orientation (crease side, dorsal side, and both side). Classification using 750-1700 and 500-1700 nm wavelength regions and dorsal side gave the highest accuracy. Light scattering, protein content, kernel hardness, starch content, and kernel color contributed to the classification of DHV and NDHV kernels.

Insect infestation downgrades the quality of grain and lowers the market value. Maghirang et al. (2003) detected the infestation by different growth stages (pupae, large larvae, medium larvae, and small larvae) of both dead and live rice weevils using a spectrometer in the wavelength range of 400-1700 nm. Calibration models, however, were developed only in the 950-1690 nm wavelength range as the visible or very near

infrared regions do not contribute in the detection of internal insect infestation (Dowell et al., 1998). They developed PLS regression models by using live insects in the calibration set and dead insect in the validation set and vice versa. The samples containing pupae or large larvae were well differentiated from sound kernels with maximum accuracy of 96%. However, samples containing small larvae were not differentiated. Their study demonstrated that calibration models can be developed from either dead or live insects which are capable of differentiating both dead and live insect infested kernels from sound grain. They found 990, 1135, 1210, 1325, 1370, 1395, 1425, 1510, 1610, and 1670 nm wavelengths as being significant. Wavelength 990 nm may be related to the loss of starch in the kernel consumed by developing insects (Ridgway et al., 1999) and wavelength 1510 nm may be related to the change in protein content of infested grains. Wavelengths 1335 and 1670 nm could be contributed to the cuticular lipids of rice weevil. The waveband at around 1425 nm was previously identified by Ridgway and Chambers (1996) and considered in a response to insect related moisture. In a similar study, Paliwal et al. (2004) evaluated the feasibility of detecting different insect life stages (i.e., eggs, larvae, pupae, and adults) of rice weevil (*S. oryzae*) and lesser grain borer (*R. dominica*) in wheat by NIR spectroscopy in the range of 400-2500 nm. The spectral data were pretreated by multiple scatter correction (MSC) and standard normal variate (SNV) and combined MSC-SNV had similar effect on the performance of classification models. Principal component analysis was used for discrimination of two species at the pupal growth stage. Two species were well separated at 25% infestations level and above. Quantitative analysis of spectral data was done by applying PLS regression to predict different level of infestations. Their PLS model gave very high prediction accuracy for

different level of infestations in three growth stages, i.e., larvae, pupae, and adults but did not predict the level of infestations in samples infected with eggs due to their small size.

Wang et al. (2003) classified healthy and fungal-damaged soybean seeds and differentiated among various types of fungal damage using NIR spectroscopy. They used PLS and ANN to develop calibration models. A two-class PLS model correctly classified more than 99% of damaged seeds. However, in five-class models, ANN gave better classification results and correctly classified healthy seeds, *Phomopsis*, *Cercospora kikuchii* Matsumoto and Tomoy, Soybean mosaic virus (SMV), and downy mildew damaged seeds with 100, 99, 84, 94, and 96% accuracy, respectively. They also transformed the NIR reflectance spectra to color space ( $L^*a^*b^*$ ) to determine the seed color of healthy and infested kernels. *Cercospora kikuchii* and SMV infested kernels had darker color than healthy seeds which was confirmed by the higher L values of healthy soybean seeds. Higher b values of the healthy soybean seeds indicated that the healthy seeds were more yellowish than fungal damaged seeds.

When fungi grow on the grain, they may produce toxins called mycotoxins and contaminate the grain. The fungus *Aspergillus flavus* Link produces lethal mycotoxins called aflatoxins. Aflatoxins can cause diseases such as acute or chronic liver disease and liver cancer. Pearson et al. (2001) studied the potential of applying NIR spectroscopic technique to detect aflatoxin-contaminated corn kernels. They used standard aflatest affinity chromatography procedure to measure the aflatoxin content in corn for reference data. Reflectance and transmittance spectra were transformed to absorbance units and absorbance ratios were computed at 15 nm interval. The best ratio features were selected and classification models were developed using PLS and discriminate analyses. The

classification model correctly classified 95% of kernels as containing either high (>100 ppb) or low (<10 ppb) levels of aflatoxin. However, the models were not able to classify the kernels in between 10 to 100 ppb aflatoxin concentration level.

Gributs and Burns (2006) predicted moisture content of wheat by developing NIR calibration models using WT and a genetic algorithm (GA). Wavelet transform uses wavelength regions (broader spectral features), instead of single wavelengths in which adjacent bands are not orthogonal (independent) due to broad absorption bands. They developed GA models using both raw as well as wavelet preprocessed data. Their calibration model gave high prediction accuracy with standard error of cross validation (SECV) of 0.096%, and standard error of prediction (SEP) of 0.27, and  $r^2$  of 0.96. However, the result using wavelet transform preprocessing did not show any significant improvement over calibration using a raw data set.

Chemical compositional analysis of cereal grain and grain products is important for their quality evaluation. Berardo et al. (2004) measured the carotenoid concentrations in maize by applying NIR spectroscopy and collecting reference values by high performance liquid chromatography (HPLC). The predicted carotenoid concentrations by the NIR model were in good agreement with those of the HPLC method ( $r^2$  value ranged from 0.82 for lutein to 0.94 for zeaxanthin). In another study, Wesley et al. (2001) developed a model to predict the protein composition (gliadin and glutenin) in wheat flour by NIR spectroscopy and compared the results with a traditional curve fitting method (chromatography). Reference gliadin and glutenin content in wheat was determined by size exclusive-high performance liquid chromatography (SE-HPLC) technique and performance of the developed PLS calibration model was compared with

curve fitting methodology. However, both of the methods did not give satisfactory results. The performance was affected by the high co-linearity of total protein content to gliadin and glutenin content. Kays et al. (2005) predicted the fat content of several intact cereal food samples by NIR spectroscopy. They collected the reference data by solvent extraction method and developed PLS calibration models. They also studied the effect of moisture and particle size on the model performance and found that inclusion of moisture and particle size variation in the calibration model improved the prediction accuracy. The prediction results showed that the model could be used for screening of many intact cereal products. The prediction capability of the quantitative NIR models is highly dependent on the reliability and precision of the reference method and reference data.

### **3.3 Applications of Digital Color Imaging**

Machine vision has found a wide applicability in agriculture and food processing industries. Brosna and Sun (2002) have thoroughly reviewed the applications of machine vision in agriculture and food industries in assessment of fruits and nuts (apple, oranges, strawberries, nuts, tomatoes, peaches and pears), vegetable inspection (mushrooms, potatoes, and others), grain classification and quality evaluation (wheat, corn, rice, barley, oats, rye), and applications in food products (pizza, bakery products, cheese, meat and meat products). Color imaging has been successfully applied for determining various quality parameters of grains (Table A.2) based on the external features such as color, texture, shape and size of the products.

Steenhoek et al. (2001) studied the potential of using color images to segment corn kernels surface damage into blue-eye mold, germ damage, sound germ, shadow in sound germ, hard starch, and soft starch. Red, blue, and green pixel values representing

the color pattern features were extracted from the images and given as input to a PNN for classification. A threshold (using a histogram plot) was applied to remove the background. Their PNN classifier did not give satisfactory results (75% accuracy) and misclassification occurred due to significant pixel overlap between similar categories. They improved the classification (95%) by grouping pixels into three categories: damage (blue-eye mold and germ damage), germ (sound germ and shadow in sound germ), and starch (soft and hard starch). This is an example of the situation where near-infrared hyperspectral imaging has the potential to classify the corn kernels into the original six classes as the compositional information from all these categories can also be extracted and used for classification.

Ni et al. (1998) developed a corn grader by applying image processing algorithms. Images of single kernels were obtained and fixed thresholding was used to convert images into binary format and kernel length, width, thickness and projected area were determined. The Ratio of the kernel thickness to width was given as input to a tree structured linear discriminant classifier to separate flat kernels from round shaped kernels. Kernel length, width, and projected area were used to classify kernels based on their size (large, medium, and small). Their system gave an average classification accuracy of 74%-90%.

Ng et al. (1998) evaluated the mechanical and mold damage in corn kernels using machine vision. The mechanically damaged corn kernels were green dyed before image acquisition so the values of G (green) component was always greater than R (red) component of the damaged portions. The G/R ratio was used as thresholding cut-off to separate damaged pixels from undamaged pixels of the corn kernels. However, fungus-

infected corn kernels were not dyed as dye adhered to only the starch portion of damaged kernels. The R (red), G (green), and B (blue) values were extracted from the images and given as input to BPNN classifier. Their model correctly classified more than 99.5% mechanical damaged kernels and 98.7% mold damaged kernels. Higher classification of mechanical damaged kernels was mainly due to contrast created by green dyeing of the damaged portion. This does not seem very practical for application at grain handling facilities.

Classification algorithms were also developed by Wan et al. (2002) using digital images to sort rice into sound, cracked, chalky, immature, dead, broken, damaged, and off-type kernels. They developed fundamental image processing and feature extraction programs in C programming language. Histogram thresholding was used to remove the background and 16 kernel features were extracted from the rice kernel images and ranked in the tabular form for each of the rice quality parameter. Logical “and” was used for the classification and a rice kernel was categorized when its extracted image features fell within the range of tabulated values of that class. The program was developed to classify on the worst fit basis, so that a kernel could be classified to the lowest quality category. This system correctly classified more than 90% kernels on average.

Zayas and Flinn (1998) detected the insects in bulk wheat samples using a machine vision system. A total of 35 image features such as mean, standard deviation, skewness, kurtosis, and several derived features were extracted from histograms of red, green, blue, hue, saturation and lightness images, for  $8 \times 8$  pixels subimages and given as input to discriminant classifiers. The developed algorithms for two-class subimage classification (insects versus non-insects) and three-class model (insects, dark grass

seeds, and others) gave 93.8% and 92.6% average accuracy, respectively. Position of the insects (ventral or dorsal side), particles clinging to the insect, dockage with similar color to insects affected the performance of the model. They also visually saw the classified images into binary image form by assigning 1(white) value to subimage pixels belonging to insects and 0 (black) value to the rest of the pixels in the subimage. Though the results in this study were satisfactory, hidden insects and early development stages were not detected in this study. However, a similar image analysis method can be applied to near-infrared hyperspectral imaging, which has the potential to detect hidden and early development stages in bulk grain as images also contains compositional information of insects.

Majumdar and Jayas (2000a, b, c, d) thoroughly studied the potential of applying image processing technique for classification of cereal grains by developing image analysis algorithms. They extracted and used texture, color, morphological features of the color images and their combinations to classify individual kernels of CWRS wheat, CWAD wheat, barley, oats, and rye. They achieved the highest average classification accuracy of 99.7 and 99.8% respectively for CWRS and CWAD when texture, color, morphological features were combined and 20 most significant features were given as input to discriminant classifier. In this study only clean grain samples were used, whereas the grain in the elevators also contains some impurities such as broken grain, other variety grain, chaff, straw, and dust. Paliwal et al. (2003) developed an algorithm to classify cereal grains (CWRS, CWAD, barley, oats, and rye) and to detect the dockage (broken wheat kernels, chaff, buckwheat, wheat spikelets, and canola seeds) present in the grains. They also extracted the morphological, color, and textural features from the

digital color images. They first selected the top 20 significant features from each of the texture, color, morphological features and then combined features were used to train a BPNN classifier. Their model correctly classified most of the cereals with well defined characteristics, however, particles with irregular and undefined features gave lower classification accuracy (90%).

Choudhary et al. (2008) used a color machine vision system to classify CWRS wheat, CWAD wheat, barley, oats, and rye. Wavelet, color, morphological, and textural features were extracted from the non-touching single kernels and classification algorithms were developed using LDA and QDA and different combination of extracted features. Linear discriminant classifier gave the highest classification accuracy (89.4 to 99.4%) by using combined wavelet, color, morphological, and textural image features as input.

### **3.4 Applications of NIR Hyperspectral Imaging**

Hyperspectral imaging has shown promising results for quality inspection of agricultural products as well as other applications (Table A.3). In a recent study, Gowen et al. (2007) reviewed the latest research conducted on application of hyperspectral imaging for quality evaluation of apple, corn, cucumber, citrus, pasta, peach, pork, potato, poultry, strawberry, cantaloupe, poultry, walnut, cherries, codfish, cucumbers, and maize.

The feasibility of hyperspectral imaging (also known as chemical imaging) was studied in pharmaceutical formulations to analyze compound distribution and quantify active pharmaceutical ingredients (API). Lewis et al. (2004) used NIR chemical (hyperspectral) images of tablets to visualize component distribution and found that three

main active ingredients *acetaminophen*, *acetylsalicylic acid*, and *caffeine* were non-uniformly distributed. They compared individual pixel spectra with the reference spectra of the each ingredient. They also developed a PLS model for trace component analysis with the help of a library of pure-component spectral data. They also compared individual pixel spectra with the reference spectra of each ingredient. Because the impurities in the tablets are of very low concentration, the spectral contribution from these components might not be easily detectable. So they subtracted the spectrum of bulk (excipient) material from each pixel in the original hypercube before model development. They counted total contaminated area (pixels) from the PLS score images and spectral library of pure components was required, which might not be available for many biological materials. In a similar work, Gendrin et al. (2007) predicted API and excipient content in two types of tablets containing a binary mixture of API and cellulose and pharmaceutical formulation with seven different compounds. They obtained the averaged spectra from the image hypercube and applied PLS, PLS2, and classical least squares (CLS) calibration technique after pretreating data with SNV and Savitzky-Golay (SG) filter. Classical least squares algorithm predicts concentration using reference spectra only, whereas PLS requires sample spectra along a full concentration range. The developed model predicted the API and excipient content (per pixel) in binary mixtures and pharmaceutical tablets. The PLS and CLS score concentration maps were developed to see the uniformity of distribution of API content in the tablets.

Tatzer et al. (2005) evaluated the potential of applying hyperspectral imaging in industrial application to identify cellulose-based materials (raw cardboard, colored cardboard, newspaper or printer paper). They used a line-scan spectrograph for scanning

the samples in the range of 900-1700 nm. A KNN based discriminant classifier achieved the highest classification accuracy of 95% using combined PCA and LDA feature reduction method.

ElMasry et al. (2007) used hyperspectral images to develop partial least squares (PLS) based calibration models to quantitatively predict moisture content, total soluble solids, and pH in strawberry. They created a binary mask to select the area of interest by removing background and then obtained averaged spectra for each hypercube. They also developed MLR prediction models by using  $\beta$  coefficients from PLS analysis to select optimum number of wavelengths. Prediction accuracy of PLS and MLR models was determined by correlation coefficient, standard error of calibration (SEC), and SEP. Both models gave good prediction results. They also did texture analysis to classify strawberry into unripe, ripe and overripe stages by cropping a 50×50 pixel area from each ripeness stage samples. The variation in the texture was associated with light intensity or gray levels. A color image for each sample was formed by extracting red (650 nm), green (500 nm), and blue (450 nm) waveband slices from a hypercube. Color images were then analyzed by gray-level co-occurrence matrix (GLCM) to determine textural features and used for classification by discriminant analysis. The highest average classification accuracy of 89.6% was achieved in identifying strawberry ripeness stage.

Kim et al. (2002) detected fecal contaminated apples using VIS and NIR reflectance imaging. They applied a simple threshold to a NIR band image to create a mask and then used the mask to remove the background. The PCA was applied to background removed images to discriminate the uncontaminated apple surface area from contaminated surface area with the help of PC score images. First they analyzed the

complete spectral region from 450 to 851 nm and then only NIR region (748 to 851 nm) to avoid the color variation effect within and between apple surfaces. The first two PC images in the NIR band showed less spatial variation within and between apple surface and thick contamination were clearly visible, however, this method could not detect very thin fecal spots.

Lu (2003) detected the bruises on the apple surface using NIR hyperspectral imaging. They applied minimum noise fraction (MNF) transform and PCA to selected region of interest (ROI) from fresh apples and bruised apples. Third MNF contained both dark and light areas but could not give the pattern information, i.e., whether dark area is bruised or light area. A new transformed image was developed by multiplication of first and third PC images. Mean pixel values from this transformed image corresponding to the dark and light areas in MNF image were calculated and compared. The areas with higher mean values were considered to be bruised surfaces. Apple variety, bruise development stage (from first to 47<sup>th</sup> day of bruise development), and spectral resolution affected the discrimination capability of the model (59-94% accuracy).

Liu et al. (2006) detected chilling injury in cucumbers by hyperspectral imaging technique using both spectral and image analysis. First they obtained the ROI mean spectra of damaged and fresh cucumbers by manually choosing the ROI from fresh and chilling injured surfaces. Reflectance difference, reflectance ratio, two-class PCA models (with pretreatments) on ROI spectra were used for classification. Two-band reflectance ratio model gave similar result to the multivariate model (PCA), and it was very simple to implement. Dual-band images (811nm/756nm) and PC images were also used for detection of injury. Dual-band image gave better discrimination than PC images. Though

more than 90% of the damaged cucumbers were correctly identified, cucumbers with insignificant injury symptom were difficult to classify.

Lawrence et al. (2003b) used hyperspectral imaging to detect fecal and ingesta contamination on poultry carcasses and compared the result with NIR spectroscopic analysis. They scanned poultry carcasses using hyperspectral imaging system (430-900 nm) and also recorded the NIR spectra of poultry carcasses by a spectrometer in the range of 400-2498 nm. They applied PCA to spectral data and plotted PCA scores and loadings to get significant wavelengths (434, 517, 565, and 628 nm) and then extracted the images corresponding to these wavelengths from the hypercube for analysis. They obtained image ratios using various combinations of the wavebands and then applied masking (thresholding) to ratio images to remove the background. Histogram stretching and Fecal thresholding were used for qualitative and quantitative analyses, respectively. They achieved 100% accuracy in a qualitative model and measured more than 96% contamination quantitatively. They used spectroscopic data to get significant wavelengths for their hyperspectral imaging system so there may be some system variation between these two instruments, which was not considered in their study. They also used only four wavebands in their analysis, which is multispectral analysis, not hyperspectral analysis. As proposed in this thesis, MVI, which uses PCA, can be directly applied to an image hypercube to get significant wavelengths and then these wavelengths can be used in the multispectral analysis in the same way as described in the above study. Use of MVI will remove the effect of the system variation and there will not be any need for additional NIR spectroscopic instruments.

Gorretta et al. (2006) investigated the potential of applying NIR hyperspectral imaging to evaluate the vitreousness of durum wheat. They scanned images of single wheat kernels and applied erosion process to remove the pixels of kernel contour. Remaining pixels were averaged to obtain the mean reflectance spectra of each kernel. Then wheat kernels were classified by applying PLS-FDA algorithm to preprocessed mean spectral data. They achieved maximum classification accuracy of 94%. The performance of the model could be improved by extending the wavelength region beyond 1100 nm as the absorption bands in NIR region (1100-1700 nm) may contribute to the performance of a classifier, however, there would be high cost associated with InGaAs detector for this wavelength region.

Cogdill et al. (2004) predicted moisture and oil content of corn kernels using NIR hyperspectral imaging. They reduced the entire hypercube data to a single transmittance spectrum by averaging the pixels of transmittance images at each of the 69 wavelengths to form NIR spectra and then applied PCR and PLS calibration techniques to predict the moisture and oil content of the corn. Optimum number of wavelengths was selected by GA. In averaged images, some information contained in the spatial variation of intensity may be lost. Principal component analysis of the averaged spectra does not account for the spatial variability of the image data as it takes the averaged value of each image slice. An alternative MVI for data reduction and selection of significant wavelength which can be directly applied to image hypercube is proposed in this thesis.

Lu (2004) used multispectral imaging for firmness and soluble solid content determination of the apples. Multispectral imaging uses only few wavelengths (2-10). He selected significant wavelengths for algorithms development by reviewing the literature.

However the significant wavelengths reported in the literature were quite different. The difference in the significant wavelengths was due to various instruments used in the measurement and data processing techniques in previous studies. The hyperspectral imaging system offers an improved method to select significant wavelengths by applying data reduction methods such as PCA to a hypercube which can then be used for multispectral analysis.

Archibald et al. (1998) used a SWNIR imaging system to classify wheat into color classes. The reference spectral characteristics of six wheat classes, three each of HRS and hard white winter (HDWW) wheat, were first obtained by NIR spectrometer and significant wavelengths were selected to predict the percent of red color of a sample by multiple linear regression. Then bulk samples of mixtures of HRS and HDWW wheat (50:50) were scanned in spectral imaging system at 11 selected wavelengths. Principal component analysis was used to analyze the data after reshaping into two-dimensional arrays. Principal component scores were mapped into pseudo images and eight PC scores were examined. The score images showed the contrast between red and white wheat; however, the performance was poorer than the spectroscopic method. The imaging approach demonstrated the effect of non-uniform illumination, and saturated and white pixels. In this study the authors did not develop any supervised classification algorithm for future classification.

Mahesh et al. (2008) investigated the feasibility of NIR hyperspectral imaging to differentiate eight western Canadian wheat classes. Bulk images of wheat samples were scanned in 960-1700 nm wavelength range at 10 nm interval (total 75 wavelengths) and an area of  $200 \times 200$  pixels around the central pixels was cropped, averaged, and

normalized using a standard 99% reflectance panel. Significant wavelengths from the averaged spectra of the wheat samples were selected by applying PROC STEPDISC (SAS, 2002) procedure using the criteria of partial least  $R^2$  and average squared canonical correlation (ASCC). Statistical discriminant analysis (linear and quadratic) and BPNN were used to develop classification algorithms. Statistical classification algorithms were developed by PROC DISCRIM (SAS, 2002) procedure based on the leave-one-out cross validation method. Two three-layer neural network architectures (standard BPNN and Wardnet BPNN) were used in developing neural net classifiers. The BPNN models correctly classified more than 90% wheat samples. Linear discriminant analysis classifier gave the best classification accuracy and correctly classified 94-100% wheat samples.

In the above study of Mahesh et al. (2008), only spectral analysis of hyperspectral data was done for feature extraction. To further improve the classification, Choudhary et al. (2009) explored the potential of the wavelet texture analysis for identification of wheat classes using the hyperspectral image data of Mahesh et al. (2008). Central area of  $256 \times 256$  pixels size at each of 75 wavelengths of a hypercube was cropped and Daubechies-4 wavelet transform was applied up to five levels of resolution. Textural features (energy and entropy) were extracted at each level in the horizontal, vertical and diagonal directions. At each resolution level an additional rotationally invariant feature was obtained by adding these three features resulting in 40 features ( $8 \times 5$ ) per slice and 3000 features ( $40 \times 75$ ) per hypespectral image. The dimensionality of this data was further reduced by PROC STEPDISC (SAS, 2002) and significant features (top 10-100) were extracted. Wheat samples were classified by statistical analysis (linear and quadratic) using PROC DISCRIM (SAS, 2002) and standard BPNN classifiers. In

another approach in this study, PCA was applied to the normalized hyperspectral data in which the first three PCs retained more than 99% variation. The PC scores images corresponding to the first three PCs were used to extract the same wavelet features resulting in 120 features (40×3). The wavelet features from each of the three score images and in combination were given as input to the statistical classifiers. The top 10-60 features from combined 120 features were also extracted and used in classification. Linear discriminant classifier discriminated more than 99% samples using the top 90 features from hyperspectral images.

Berman et al. (2007) used hyperspectral imaging to classify wheat as sound or stained (discolored) grains. Hyperspectral images were generated by pixels based (point based scanning) spectrometer in the 350-2500 nm range, at 1 nm interval, producing 2151 reflectance values per pixel spectra which took nearly 10 h of scanning time for 300 kernels. The first 70 wavelengths were removed due to excessive noise. The data size was further reduced to 208 from 2151 by co-averaging 10 consecutive wave bands. The spectral data were analyzed using 420-2500 nm (all 208 spectral bands), 420-1000 nm (58 spectral bands) and 420-700 nm (28 spectral bands) spectral range. Penalized discriminant analysis (PDA) was used to classify wheat as sound or stained. More than 95% correct classification was achieved and classification results were similar in both full spectral range as well as reduced spectral ranges. Instead of using all wavelengths, significant wavelengths can be selected and used for classification which will significantly reduce the scanning and classification time.

Shahin and Symons (2008) investigated the potential of NIR hyperspectral imaging to classify wheat into vitreous and non-vitreous kernels. Bulk wheat samples were

scanned in 950-2450 nm wavelength range and reflectance spectra were obtained by averaging kernel pixels. Spectral data showed clear spectral differences between vitreous, starchy, piebald, and bleached kernels after pre-treating with SG smoothing and second derivative. Their study demonstrated the high ability of hyperspectral imaging to develop supervised classification algorithms for classification of vitreous wheat kernels.

Gowen et al. (2009) used hyperspectral imaging to identify freeze damaged mushroom samples. They used pushbroom type imaging system in scanning of mushroom sample in 400-1000 nm wavelength range. Mean reflectance spectra were obtained from cropped areas in the central portion of the mushroom. Spectral data were pre-treated by applying SNV and dimensionality of the data was reduced by PCA. Classification algorithm was developed using LDA to classify mushrooms into healthy and frozen classes and 100% healthy and more than 95% frozen samples were correctly classified.

### **3.5 Summary of the Literature Reviewed**

Most of the above discussed analytical methods currently in use for detection of insect and fungal damage, and sprouted kernels, have several disadvantages as they are slow, labor intensive, inconsistent, destructive, and subjective in nature. Near-infrared spectroscopy and image processing have demonstrated many applications in grain quality analysis. However, from the discussion of the literature review, it is clear that hyperspectral imaging takes the advantage of both spectroscopic and image processing techniques. Hyperspectral images can be simultaneously analyzed to see the spatial variations in an image slice as well as variation in spectral response of (across the wavelength) a pixel. Chemometric methods are used for qualitative and quantitative

model development using spectral data. Principal component score images and concentration maps clearly show the compositional distribution (spatial variation) in an image. The choice of selection of appropriate wavelength range, preprocessing, and analytical method depends on the type of analyte and the objective of analysis. Principal component analysis is the most commonly used method for data reduction. Multivariate image analysis, which is a PCA based data reduction technique, can be used to reduce the dimensionality of hyperpsectral data and select significant wavelengths. These significant wavelengths can be used in multispectral analysis in which processing speed is reduced by several folds due to use of only few selected wavelengths. Multispectral imaging system uses 2-10 wavelengths whereas hyperspectral imaging system uses more than 100 wavelengths. In some NIR spectroscopy applications, combining features from VIS wavelength region with the NIR region improved the classification performance compared to classification using VIS or NIR region alone. So combining color image features (RGB) with NIR image features might improve the classification performance of the model developed using these features. Supervised classification models can be developed using simple statistical classifiers, e.g., LDA, QDA, and Mahalanobis classifiers. In LWNIR hyperspectral data, the performance of LDA was better than BPNN, which could be possibly due to linear response of InGaAs detector. The BPNN is likely to give better classification performance than statistical classifiers if the input training data is non-linear.

## 4. MATERIALS AND METHODS

The discussion in this chapter is mainly focused on sample preparation, data collection, and algorithms development.

### 4.1 Sample Preparation

**4.1.1 Insect-infested kernels** Canada Western Red Spring wheat (cv. AC Barrie) samples at approximately 15% moisture content (wet basis) were used in this study. Moisture content was determined using a standard oven method by drying about 10g samples in triplicate at 130 °C for 19 h (ASAE, 2003). Healthy wheat samples and wheat samples damaged by the insect species rice weevil (*Sitophilus oryzae* (L.)), lesser grain borer (*Rhyzopertha dominica* (F.)), rusty grain beetle (*Cryptolestes ferrugineu* (Stephens)), and red flour beetle (*Tribolium castaneum* (Herbst)) were prepared at the Cereal Research Centre, Agriculture and Agri-Food Canada, Winnipeg, Canada. Adult insects of each species were mixed with 50 g whole wheat and kept for 4 weeks at 30 °C temperature and 70% relative humidity. Visually damaged kernels were selected and stored at the same temperature and relative humidity conditions prior to imaging. A total of 300 healthy kernels and 300 kernels damaged by each insect species were selected for imaging. Each sample was divided into an independent training set (80% of the kernels) and a test set (20% of the kernels) (Singh et al., 2009a).

**4.1.2 Fungal-infected kernels** The species of storage fungi namely *Penicillium spp.*, *Aspergillus glaucus*, and *Aspergillus niger* Van Tieghem were taken from old wheat at the Cereal Research Centre, Winnipeg, Canada. The infected wheat seeds were plated on filter paper saturated with 7.5 ml aqueous NaCl in petri dishes. After 7 days, pure fungal lines from infected seeds were placed on potato dextrose agar for one week at 30 °C.

Then agar plus mold were placed in 200 ml sterilized water in a plastic spray bottle with 1 drop of Tween 20 and shaken up. A 20 kg sample of CWRS wheat (cv. AC Barrie) was moisturized to 17% moisture content (wet basis). Wheat grains were surface sterilized by soaking in 1% sodium hypochlorite for 2 min. The wheat grains were thoroughly rinsed in sterilized water and placed on a paper towel for 2 h. For each fungus approximately 1 kg moistened wheat was placed in large plastic tubs, a few layers deep and misted in a fume hood and covered with loose plastic bags at 30 °C for 4 wk. Then each sample was shaken in a fume hood for 1 min in four plastic bags sequentially and poured into a fifth plastic bag before imaging. This process simulated the mixing of grain during handling. The control samples had moisture of 15% (wet basis) (Singh et al., 2007).

**4.1.3 Artificially-sprouted and midge-damaged kernels** Healthy, artificially-sprouted, and naturally midge-damaged wheat samples were used in the analysis. Sprouted wheat kernels were created for CWRS wheat (cv. AC Barrie). A 4 kg sample was first surface sterilized by soaking into 2.5% sodium hypochlorite aqueous solution for approximately 10 min. The wheat sample was thoroughly rinsed with distilled water and soaked in excess distilled water overnight. The sample was then spread on paper towels for about 48 h to germinate. The germinated sample was stored in a freezer at -40 °C and freeze-dried to 14% moisture content (wet basis). Sprouts were physically removed from the wheat kernels prior to the imaging and there was no visible damage except some swelling. The falling number of the sprouted and healthy samples was determined in the laboratory (Approved method 56-81B, AACC International, 2000). The falling number test is used to determine the levels of alpha-amylase enzyme. A high falling number value indicates low alpha-amylase level. The falling number of sprouted samples was 61

which indicated very high level of sprouting and the falling number of the healthy samples was 343. Three hundred healthy and 300 artificially sprouted kernels were randomly picked for imaging.

Midge-damaged composite wheat (CWRS) samples from five different growing regions namely Camrose East, AB (4.3% of the kernels midge-damaged), Vegreville, AB (5.2% midge-damaged), North Battleford, SK (5.5% midge-damaged), Cutknife, SK (4.1% midge-damaged), and Yorkton, SK, (6.2% midge-damaged) across western Canada were procured from Cargill Foods. The falling numbers of these samples were 334, 271, 387, 325, and 190, respectively. Interestingly, one of the midge-damaged samples showed higher falling number value (387) than the healthy kernels (343). These falling numbers represent bulk samples from which obvious midge-damaged kernels were selected. Three hundred midge-damaged kernels from each of the five locations were manually selected by using a microscope and were verified by an expert on midge-damage at the Cereal Research Centre, Winnipeg. For all samples, the 300 kernels were divided into independent training (80%) and test sets (20%) and used in classification (Singh et al., 2009b).

## **4.2 Imaging Systems**

**4.2.1 Long-wave NIR hyperspectral imaging system** The LWNIR hyperspectral imaging system reported by Singh et al. (2007, 2009a, b) was used in this study (Fig 4.1). The imaging system consisted of a thermoelectrically cooled InGaAs camera (Model no. SU640-1.7RT-D, Sensors Unlimited Inc., Princeton, NJ, USA) of 640×480 pixels detector size and 27 µm pitch, a 25 mm F1.4 C-mount lens (Electrophysics Corp., Fairfield, NJ, USA), and two VariSpec LCTFs (Model no. MIR06, Cambridge Research

and Instrumentation Inc., Woburn, MA, USA) giving a tuning range of 900-1700 nm. The minimum resolution of the system was 0.01 nm which was limited by the internal electronics controlling of the LCTFs. Two 300 W halogen-tungsten lamps (Ushio Lighting Inc., Cypress, CA, USA) fixed at 45° angles and 0.5 m away from the imaging area were used as illumination sources. The hyperspectral data were acquired through a Dell Optiplex GX280 Intel(R) computer (Dell Inc., Round Rock, TX, USA) using a digital data acquisition frame grabber board (Model no. NI PCI-1422, National Instruments Corp., Austin, TX, USA) compatible with RS-422 signals generated from the InGaAs camera. System controls developed in LabVIEW (Version 1, National Instruments, Austin, TX, USA) were used to align the imaging system, acquire images, and store hyperspectral image data in 12-bit binary file. The program stores the imaging system setup, wavelength range and number of wavelength slices along with the hyperspectral imaging data in a file.

**4.2.1.1 Image acquisition** Three hundred kernels from a sample were randomly picked, scanned with the same orientations, and then the next sample was scanned in a similar way. For every image, five non-touching randomly selected kernels were manually placed on a black plastic board with the same orientation in the imaging area of the hyperspectral imaging system and scanned at 60 evenly spaced wavelengths in the 1000-1600 nm range (10.17 nm intervals). Initially (Singh et al., 2007), 20 uniformly distributed wavelengths were selected in the 1000-1600 nm range. The five kernels were imaged at a time to get uniform intensity of light and the best resolution as well as to reduce the time of imaging and storage space for single kernels. The scanned hyperspectral images (hypercube) were stored with an *.spc* extension in their file name.

The imaging of each sample (60 wavelengths) took approximately 2 min. The wavelength region 1000-1600 nm was selected due to the high quantum efficiency of the InGaAs detector (>70%) in this region. All the images were acquired at the relatively high integration time of 250 ms to improve the signal-to-noise ratio. Prior to image acquisition, the imaging system was aligned at the central wavelength (1300 nm) of the spectral range used and a dark-current image was acquired by blocking the entrance of the camera. The system was stabilized by switching on the sensor for approximately 1 h before image acquisition. Dark-current images were acquired at the beginning of each imaging session which was typically 2-4 h long. The LabVIEW control program automatically subtracted the dark-current image from subsequent acquired images. The image data were transformed into reflectance using a standard Spectralon reflectance panel of known reflectance (99%) (Labsphere Inc., North Sutton, NH, USA).

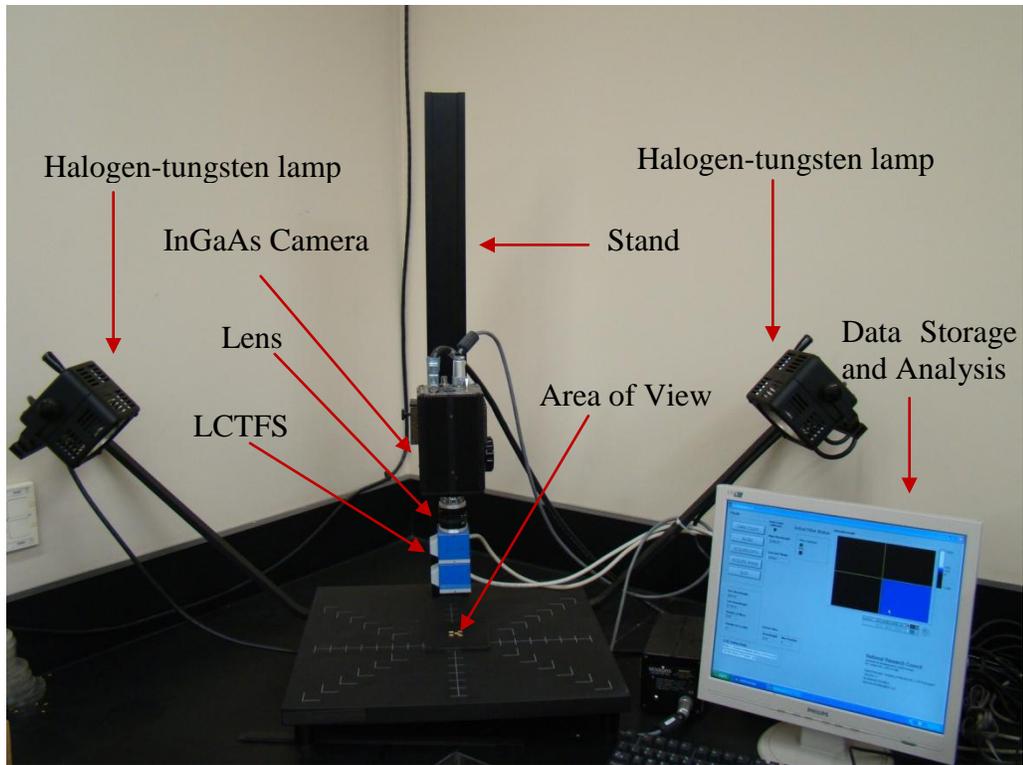
**4.2.1.2 Data analysis** A program was written in MATLAB (Mathworks Inc., Natick, MA, USA) to import the image files, and to display and analyze the hyperspectral data. The insensitive pixels in the images were removed by applying a 3×3 median filter. Most of the cameras have a few insensitive pixels due to the sensor defect, which are called dead pixels. The image slices at each wavelength were automatically co-added 10 times during image acquisition to improve the signal-to-noise ratio. Single-kernel images from 5 non-touching kernels in the original images were labelled and obtained by implementing a code developed in MATLAB (Singh et al., 2007). The original hypercube data were converted from matrix format to grayscale intensity images. Kernels in the hyperspectral image were labeled (1-5) by the *bwlabel* MATLAB function after converting an image slice to a binary image using automatic thresholding. The function

*bwlabel* returned a matrix of the size of the original image but labeled all the connected objects in the binary image, excluding the background. Image erosion was used in binary images to account for useful pixels in the kernels prior to labeling. An NIR image (a) and five labeled kernels after removing background (b) are shown in Fig. 4.2. A MVI program written in MATLAB was used to analyze each of the five labeled kernels in the image. The MVI was performed using PCA (Geladi and Grahn, 1996). A simple multivariate image had two pixel coordinates (width and height) and a variable index (wavelength value) making a three-dimensional array (hypercube). The hypercube data were reshaped into a two-dimensional array by rearranging all the pixel intensities (reflectance) of a kernel into a column at each of the 60 wavelengths. This resulted in a  $k \times 60$  sized two-dimensional array, where  $k$  is the total number of pixels in a labelled kernel. Principal component analysis was then applied to the reshaped two-dimensional data set of each kernel.

The most significant wavelengths corresponding to the highest factor loadings of the first principal components (PC) (second PC was also investigated) were selected and used for feature extraction. In the case of single kernel analysis, the discriminant analysis cannot be directly applied to the PCA reduced data set, e.g., PCA score due to the variation in kernel size and thus total number of pixels. Discriminant classifiers require matrices having the same number of elements in each row (each row corresponds to a sample image). However, this method can be used in bulk grain analysis by automatically selecting a ROI from the bulk grain images, applying MVI and then applying discriminant analysis. Therefore, for single kernel analysis a different approach was used to select the image features with significant discrimination capability. In the first study

(Singh et al., 2007) statistical features namely maximum, minimum, mean from the images corresponding to the significant wavelengths were extracted and used in classification development. Then three more statistical features namely median, standard deviation, and variance were extracted and added (Singh et al., 2009b). Ten histogram features from the transformed (reflectance) images at each significant wavelength were also extracted by binning the reflectance into 10 equally distributed reflectance groups between 0 and 1. These features were combined with statistical features and used as input in various classifiers (Singh et al., 2009a).

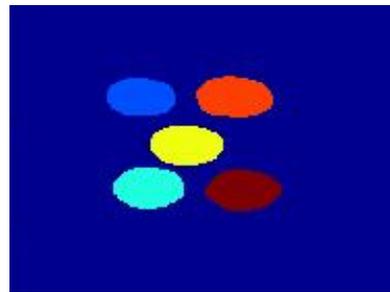
**4.2.2 Short-wave NIR hyperspectral imaging system** The imaging system consisted of a  $532 \times 256$  pixel size FFT-CCD area-scan image sensor (Model no. C7042, Hamamatsu Photonics, Hamamatsu, Japan) working in VIS and SWNIR with two stage thermoelectric cooling and two electronically tunable LCTF filter devices (Model no. MIR06, Cambridge Research and Instrumentation Inc., Woburn, MA, USA) for rapid wavelength selection in VIS and SWNIR regions. The SWNIR imaging system is shown in Fig. 4.3. The camera was attached to a shutter and a lens was attached to the other end of the shutter. A sliding unit was used to move the appropriate filter in the optical path of the lens. A slider was screwed below the lens in such a way that it moves the selected filter into the exact optical path of the lens. The user had the option to select VIS, SWNIR, and hard filter through the user interface. The lens and filter slider moved independently and the lens could be used to adjust the focus. The whole unit was mounted in a copy stand with flexibility of vertical movement to adjust the focus and



**Fig. 4.1** Long-wave near-infrared hyperspectral imaging system

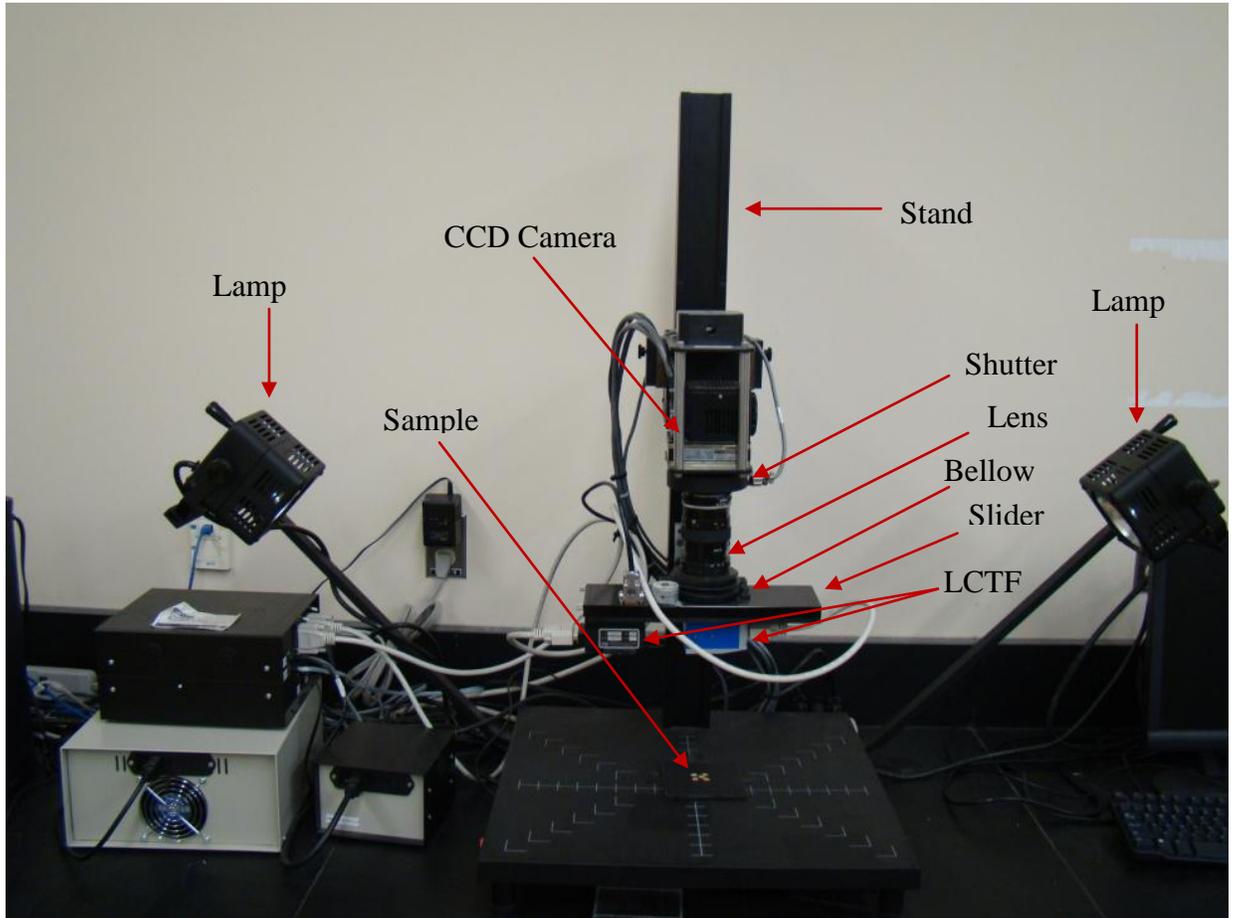


**(a)**



**(b)**

**Fig. 4.2** (a) NIR image and (b) labeled kernels in image (a) separated from the background



**Fig. 4.3 Short-wave NIR hyperspectral imaging system**

alignment. Two 300 W halogen-tungsten lamps fixed at 45° angles and 0.5 m away from the imaging area controlled by a dimmer were used as illumination sources. The image data were acquired by a data acquisition board (Model no. NI PCI-1422, National Instruments Corp., Austin, TX, USA) operating at 333K samples/s at 16-bit. System controls developed in LabVIEW (Version 1, National Instruments, Austin, TX, USA) were used to align the imaging system, control the light intensity, record dark current, acquire images, and store hyperspectral image data in 12-bits binary file. The program stores the imaging system setup, wavelength range and number of wavelength slices along with the hyperspectral imaging data in a file.

**4.2.2.1 Image acquisition** Similar to the previous section, for every image, five non-touching randomly selected kernels were manually placed on a black electric tape wrapped in black paper board with the same orientation in the imaging area of the hyperspectral imaging system and scanned at 41 evenly spaced wavelengths in the 700 to 1100 nm range (10 nm intervals). The kernels were kept on electric tape to restrict their movement while moving the samples from the SWNIR system to the color imaging system during scanning and to retain the kernel label information while shifting the kernels. The imaging system was switched on for about 1 h prior to imaging to reach to stabilized condition. The SWNIR filter was selected for imaging and light intensity was set to 60% of the maximum to avoid saturation. The imaging system was aligned at the 880 nm wavelength. Dark-current images were acquired at the beginning of each imaging session which was typically 2-4 h long. The control program automatically subtracted the dark-current image from subsequent acquired images. The image data were transformed

into reflectance using a standard Spectralon reflectance panel of known reflectance (99%) (Labsphere Inc, North Sutton, NH, USA).

**4.2.2.2 Data analysis** A program, similar to the one described in previous section, was written in MATLAB (Mathworks Inc., Natick, MA, USA) to import the image files, and to display and analyze the hyperspectral data.

A total of six statistical image features, namely, maximum, minimum, mean, median, standard deviation, and variance from the images corresponding to the significant wavelengths were extracted. Histogram features from the transformed (reflectance) images at each significant wavelength were also extracted by binning the reflectance into equally distributed reflectance groups between 0 and 1 at 0.05 intervals. These features were used as input in various discriminant and BPNN classifiers.

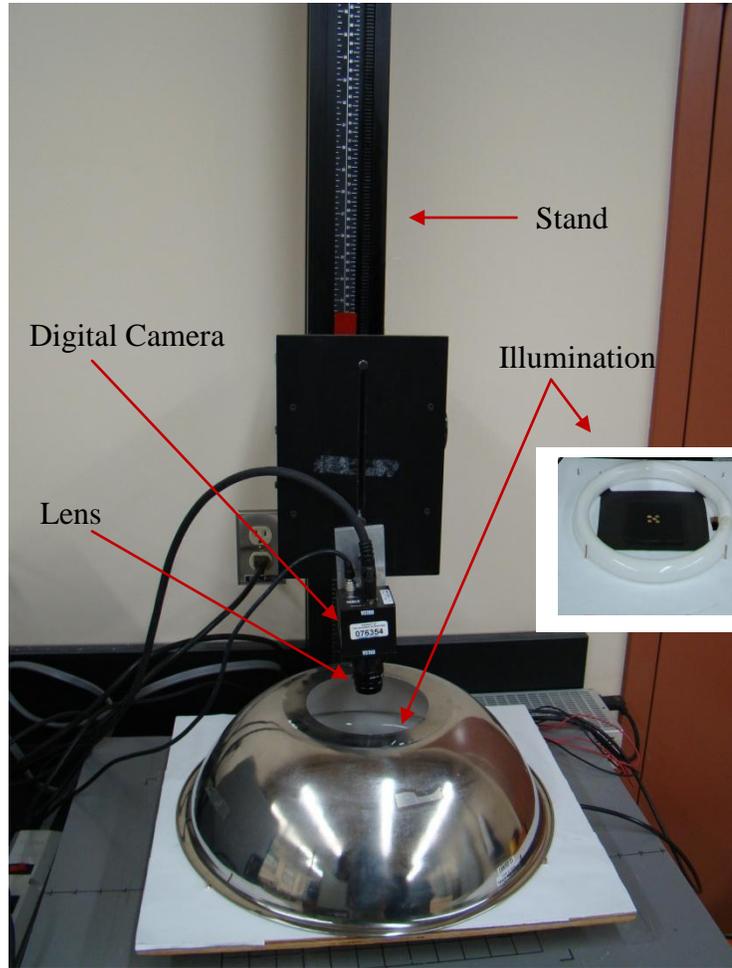
**4.2.3 Color imaging system** The imaging system available at the Canadian Wheat Board Centre for Grain Storage Research was used to scan the healthy and damaged wheat samples (Fig. 4.4). The digital imaging system mainly consisted of a high performance 2-megapixel interline transfer CCD  $1600 \times 1200$  resolution camera (Model no. DS-22-02M30, DALSA Corporation, Toronto, ON), a zoom lens of 10 to 120 mm focal length (fixed in front of camera), and a camera stand (Model no. m3, Bencher Inc., Chicago, IL) to support the camera. The camera used a standard Camera Link connector to connect to a frame grabber board (Matrox Meteor-II PCI frame grabber, Matrox Electronic Systems Ltd, Montreal, PQ) attached to a Pentium 4.3 GHz computer. Illumination to the camera was provided by a 32 W fluorescent 305 mm diameter circular light source (FC12T9/CW, Philips, Singapore). The system was calibrated using a standard gray card (CPM Inc., Dallas, TX). Three hundred randomly selected kernels from each sample

were scanned. Five non-touching kernels were placed at a time in the viewing area of the camera and images were acquired and saved in the computer for further analysis.

**4.2.3.1 Image acquisition** Five non-touching kernels placed on the electric tape were moved to the color imaging system immediately after data acquisition in the SWNIR imaging system and images were acquired and saved as *tif* files. The movement of the kernels on the black paper board while shifting from the SWNIR system to the nearby color area-scan camera was restricted to label (identify) the same kernel in both the systems as any movement in kernels will change the order of counting and labeling, making the addition of features from both systems difficult and inaccurate.

**4.2.3.2 Data analysis** The color image features were extracted by using the machine vision algorithms developed by the researchers of the Canadian Wheat Board Centre for Grain Storage Research. The actual size of each acquired image was  $800 \times 600$  pixels. An image area of  $351 \times 251$  pixel size containing all five kernels was cropped by a code developed in MATLAB (Mathworks Inc., Natick, MA) using a moving window of  $351 \times 251$  and saved without any image compression or loss of data. A total of 230 features (123 color, 56 textural, and 51 morphological features) from each kernel of the five non-touching kernels of the cropped images were extracted.

Color features extracted from the grain kernels were: mean, variance, ranges, moments (1-4), and histograms (1-32) of the red (R), green (G), and blue (B) color and mean, intensity, and range of hue (H), saturation (S), and intensity (I) values. Textural features were derived from the GLCM and gray level run length matrix (GLRM). The GLCM gives information about the distribution of gray level intensities and GLRM contains



**Fig. 4.4 Color imaging system**

information about the occurrence of collinear and consecutive pixels of similar gray levels in an object. The image features derived from GLCM were: mean, variance, uniformity, correlation, cluster shade, entropy, homogeneity, and inertia from the gray, red, blue, and green bands. The image features extracted from GLRM were: short run, long run, color non-uniformity, run length non-uniformity, entropy, and run percent from the gray, red, blue and green bands. The morphological features extracted from grain kernels were: area, perimeter, maximum radius, minimum radius, mean radius, major axis length, minor axis length, shape moments (1-4), radial Fourier descriptors (1-20), and boundary Fourier descriptors (1-20). The details of these features are given in Paliwal (2002), Visen (2002), and Karunakaran (2002).

Since the extracted feature set was very large and some features might be redundant, features were selected after ranking in order of their significance using STEPDISC procedure in SAS (Version 9.1, SAS Institute Inc., Cary, NC) and the average squared canonical correlation as the criteria of significance. The STEPDISC procedure was applied to the two data sets containing only 123 color features and 230 combined morphological, color, and textural features. Top ten features from each data set were selected as the input for development of statistical discriminant classifiers as initial results using more features (20 features) did not significantly improve the classification accuracy. Performance of the BPNN was also investigated using these data sets as input.

### **4.3 Model Development**

The classification algorithms were developed by various statistical discriminant analysis classifiers (linear, quadratic, and Mahalanobis). Linear discriminant analysis classifier uses pooled co-variance in Bayes' criteria to assign an unknown sample to one

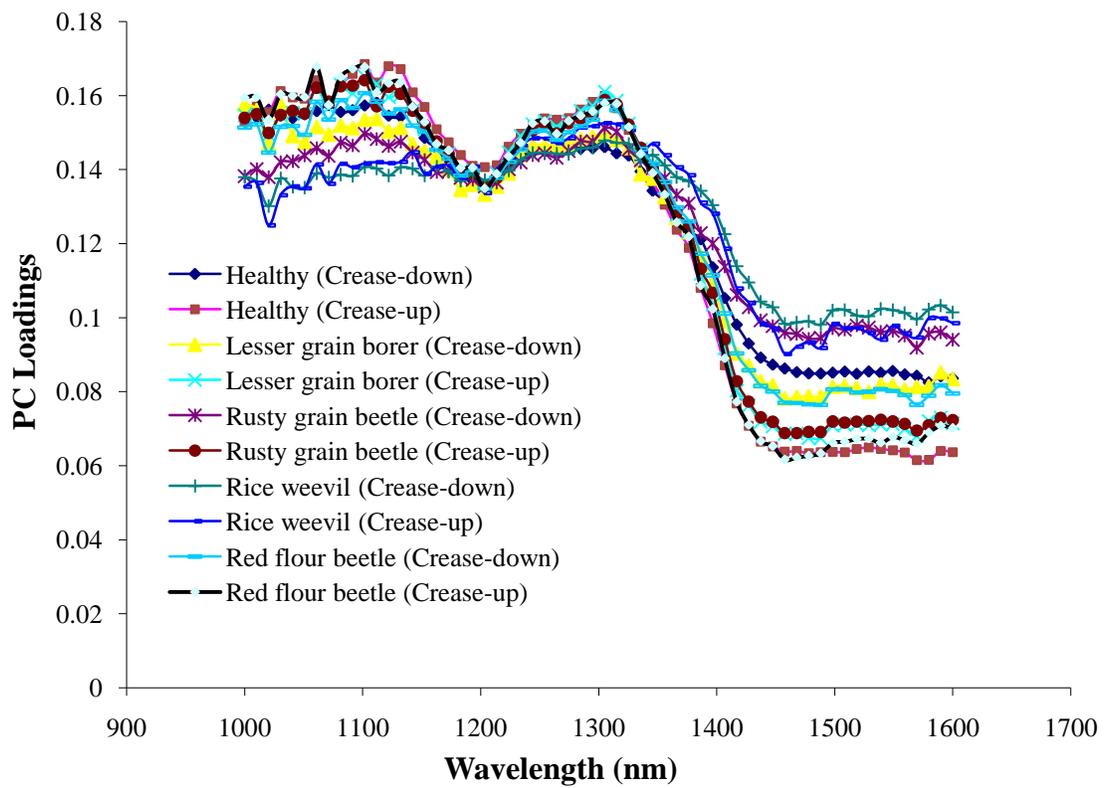
of the predefined groups (Naes et al., 2002). In many biological applications, data do not follow normal distribution, so using pooled co-variance may not give appropriate classification. Quadratic discriminant analysis classifier uses covariance of each class instead of pooling them in Bayes' criteria for grouping of unknown samples. If the covariance of the classes is equal, both LDA and QDA give the same results. The Mahalanobis discriminant classifier is a simplified form of the LDA with assumptions that all the known groups have equal posterior probabilities. This method gives good classification when the data distribution is elliptically concentrated and an unknown sample is assigned to the group with minimum Mahalanobis distance.

Back propagation neural network was also investigated for classification algorithm development by SWNIR and color imaging systems. A three layer (input, hidden, and output) back propagation network was developed in MATLAB. The hidden layer with three nodes had *tansig* transfer function. The output layer had linear transfer function. The epoch was set to 1000. The network was trained using *trainbr* function which uses automated regularization (Bayesian) criteria to improve the generalization performance. A number of other BPNN training functions in MATLAB were also investigated; however, *trainbr* function gave better classification. Changing the number of nodes in hidden layer also did not significantly improve the performance of classifiers. In function approximation and small sample set, Bayesian regularization gives better performance compared to early stopping criteria and there is no need of independent validation set. Each sample of three hundred kernels was divided into independent training (80%) and test set (20%).

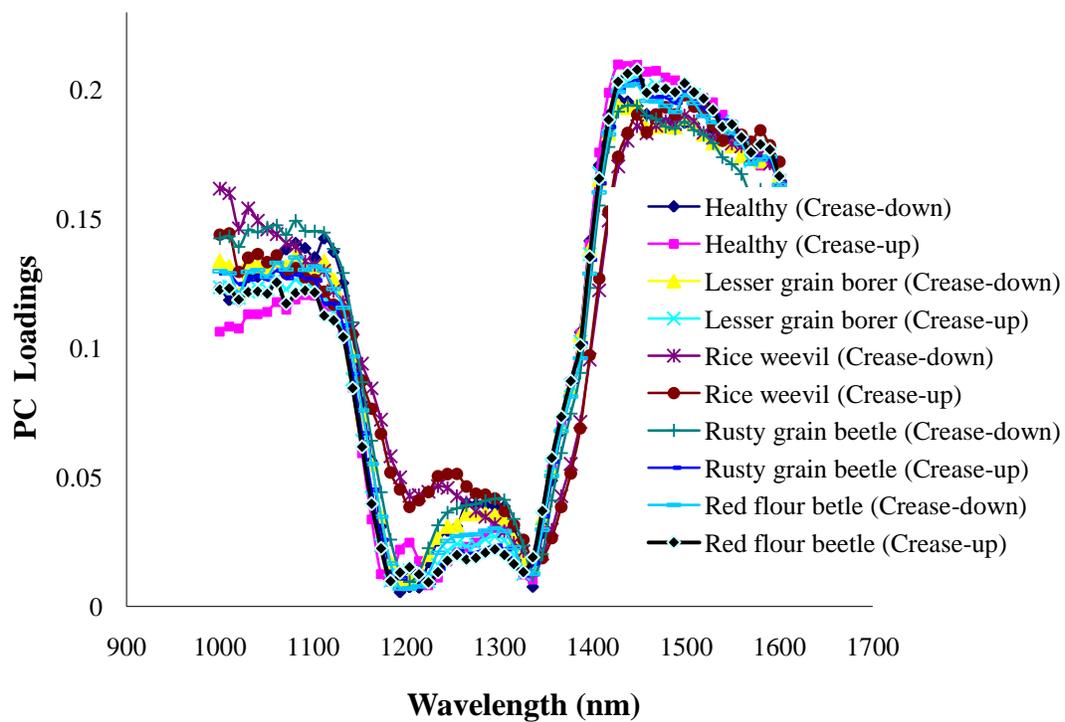
## 5. RESULTS AND DISCUSSION

### 5.1 Insect Damage Detection

**5.1.1 Detection of insect-damaged kernels using LWNIR hyperspectral imaging** First and second PC factor loadings which accounted for nearly 94% and 5% variability, respectively, in the hyperspectral data were used for wavelength selection (Singh et al., 2009a). Two wavelengths at 1101.69 and 1305.05 nm corresponding to the highest factor loadings of the first PC (Fig. 5.1) and another wavelength at 1447.46 nm corresponding to the highest factor loading of the second PC (Fig. 5.2) were selected as significant wavelengths for feature extraction. A total of 48 features (six statistical features and 10 histogram features at each wavelength) were extracted and used in the classification. The wavelength region of 1100 to 1300 nm corresponds to the C-H 1<sup>st</sup> and 2<sup>nd</sup> overtones and C-H combination band and the significance of wavelengths in this region can be associated with absorption by starch molecules. Wavelengths 1135 and 1325 nm were found to be significant by Maghirang et al. (2003) for insect detection in wheat. Dowell et al. (1999) related the peak at 1330 nm to the cuticular lipids present in *S. oryzae*. Wavelength regions of 1130-1200 and 1300-1400 nm were found to be significant by Baker et al. (1999) in identifying parasitized *S. oryzae* in wheat. Wavelengths of 1420 nm (Dowell et al., 1999) and 1440 nm (Toews et al., 2007) were selected as significant for insect detection. Insect-damaged wheat kernels have less starch compared to healthy kernels, due to consumption of starch by insects during their development. This is also evident from the pseudo-color images obtained by mapping the first PC scores of healthy and insect-damaged kernels (Fig. 5.3). The healthy kernels (germ facing left and crease-



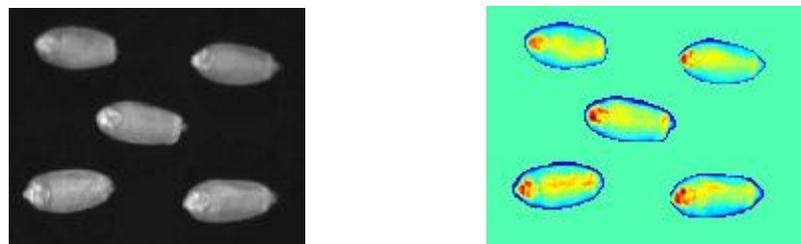
**Fig. 5.1 First principal component (PC) mean factor loadings of healthy and insect-damaged kernels (Long-wave near-infrared)**



**Fig. 5.2 Second principal component (PC) mean factor loadings (absolute values) of healthy and insect damaged kernels (Long-wave near-infrared)**

down in Fig. 5.3) show the dark germ region, but in most of the damaged kernels the germ region is not highlighted and the darker region is the endosperm area. The actual grayscale images (Fig. 5.3) do not show any clear differences between healthy and damaged kernels.

Use of all 48 features did not classify the grain due to the formation of non-positive definite covariance matrices in the discriminant analysis. The cause of this non-positive covariance matrix might be due to the linear dependency of one or more input features on other features. To overcome this problem, the input features were divided into statistical image features ( $6 \times 3 = 18$ ) and histogram image features ( $10 \times 3 = 30$ ). The same problem was observed using histogram features for classification which indicated that the histogram features have a near perfect linear dependency. Thus, only seven histogram features (groups between 0.1 to 0.8 reflectance) from the images at 1101.69 nm corresponding to the highest factor loadings of the first PC were considered as histogram features and used in classification. The same number of histogram features at the other two significant wavelengths gave lower classification accuracy. The classification models were developed by using statistical image features from significant wavelengths of the first PC (1101.69 and 1305.05 nm), combining statistical image features of first PC wavelengths (1101.69 and 1305.05 nm) with second PC (1447.46 nm), histogram features, and combining histogram features with the first PC statistical image features. Using statistical features from only one of the significant wavelengths gave less accurate classification results. The classification algorithms were developed utilizing the LDA, QDA, and Mahalanobis discriminant classifiers. The Mahalanobis classifier, however,



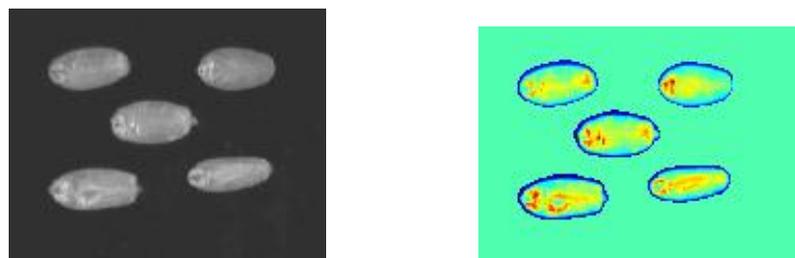
Healthy kernels



Rice weevil



Lesser grain borer



Rusty grain beetle



Red flour beetle

**Fig. 5.3 Images of healthy and insect damaged kernels (NIR images (1305.05 nm) on left and first PC score images on right)**

gave very poor and inconsistent classification accuracy so the results are not reported here.

Binary classification models were developed to differentiate wheat kernels damaged by each insect species from healthy undamaged kernels. A total of 600 kernels (300 healthy and 300 insect-damaged kernels for each insect species) in both crease-up and crease-down orientation were used for binary classification. The results of binary classification using the statistical features (12 features) of the images selected by the first PC (1101.69 and 1305.05 nm) are given in Table 5.1. The linear discriminant classifier accurately discriminated 86.7 to 100% healthy and 86.7 to 98.3% insect-damaged crease-down kernels. The classification accuracy of crease-up samples was slightly lower, especially in discriminating the healthy kernels from the *T. castaneum*-damaged kernels. The quadratic discriminant classifier also produced a similar classification. The classification results of binary classification using statistical features (18 features) of images selected by the first and second PC (1101.69, 1305.05, and 1447.46 nm) are given in Table 5.2. The linear discriminant classifier correctly differentiated 96.7-100% healthy and 95-100% insect-damaged creased-down samples. The quadratic discriminant classifier accurately discriminated 88.3-98.3% healthy and 96.7-100% insect-damaged crease-down samples. The lowest classification was found in discriminating *T. castaneum* crease-up samples with 73.3% and 78.3 % accuracy using first PC features and second PC features, respectively. *Tribolium castaneum* beetles consume most of the germ and occasionally endosperm surrounding the germ (White and Lambkin, 1988) causing less visible damage compared to other insects.

**Table 5.1 Binary classification (20% test set) using statistical features (1101.69 and 1305.05 nm wavelengths) for healthy wheat kernels and those damaged by various insect species**

Sample (n=300)	Classification by discriminant analysis (%)			
	Linear		Quadratic	
	Crease-down	Crease-up	Crease-down	Crease-up
Healthy	100.0	100.0	95.0	98.3
<i>Sitophilus oryzae</i>	98.3	100.0	100.0	100.0
Healthy	98.3	83.3	88.3	88.3
<i>Rhizopertha dominica</i>	96.7	93.3	95.00	100.0
Healthy	100.0	96.7	100.0	80.0
<i>Cryptolestes ferrugineus</i>	95.0	98.3	88.3	96.7
Healthy	86.7	70.0	86.7	73.3
<i>Tribolium castaneum</i>	86.7	71.7	90.0	100.0

**Table 5.2 Binary classification (20% test set) using statistical features (1101.69, 1305.05, and 1447.46 nm wavelengths) for healthy wheat kernels and those damaged by various insect species**

Sample (n=300)	Classification by discriminant analysis (%)			
	Linear		Quadratic	
	Crease-down	Crease-up	Crease-down	Crease-up
Healthy	100.0	100.0	93.3	98.3
<i>Sitophilus oryzae</i>	98.3	96.7	100.0	100.0
Healthy	96.7	90.0	88.3	88.3
<i>Rhizopertha dominica</i>	100	91.7	100	100.0
Healthy	98.3	100.0	98.3	93.3
<i>Cryptolestes ferrugineus</i>	96.7	93.3	95.0	98.3
Healthy	96.7	100.0	88.3	93.3
<i>Tribolium castaneum</i>	95.0	73.3	96.7	78.3

The binary classification models were also developed to discriminate healthy kernels from damaged kernels using histogram features (Table 5.3). Both LDA and QDA correctly classified 90-100% healthy and damaged kernels in both crease-up and crease-down orientations. The lowest classification accuracy, for *T. castaneum*, also improved from 73.3% (Table 5.2) to above 90%. This improvement in classification might be due to the capability of certain histogram groups in damaged kernels to capture the discriminative information for classification.

Finally, a binary classification model was developed by combining statistical features from the first PC (1101.69 and 1305.05 nm wavelengths) and histogram features (Table 5.4). The linear discriminant classifier discriminated all the healthy kernels and 98.3-100% of the damaged samples except the *T. castaneum* crease-up samples (88.3% accuracy). The QDA classifier differentiated 90-100% crease-down and 96.7-100% crease-up healthy and insect-damaged kernels. So combining the statistical image features from the first PC wavelengths with histogram features gave the highest and most consistent classification accuracies among all four binary classifiers.

Binary classification models gave very high classification accuracies, but the real-time wheat samples are likely to contain kernels infested with more than one insect species. So we developed two-way classification models by grouping the healthy kernels in one group and kernels damaged by all four insect species into another group. The two-way classification accuracy of discriminant classifiers using statistical features from the first PC wavelengths (1101.69 and 1305.05 nm) are given in Table 5.5. Both the LDA and QDA correctly classified 85-100% of the healthy kernels. The classification accuracy of damaged kernels was also very high except for the kernels damaged by *T. castaneum*

(61.7% accuracy). To improve the classification accuracy, a two-way classification model was developed by combining the statistical features from the first PC wavelengths (1101.69 and 1305.05 nm) with the second PC wavelength (1447.46 nm) (Table 5. 6). Adding the features from the second PC wavelength did not significantly improve the classification of kernels damaged by *T. castaneum* (65.0%), which indicated that the second PC did not hold the information for detection of insect-damaged kernels. The wavelength 1447.46 nm selected by the loadings of the second PC falls in the moisture absorption region (1350-1500 nm) (Murray and Williams, 1990) so the second PC might have information to explain the variation in the moisture in the grain and not much information for detection of insects.

Given the discrimination capability of histogram features in binary classification, a two-way classification model was also developed using histogram features to classify healthy and damaged kernels (Table 5.7). All the healthy and damaged kernels in both crease-down and crease-up orientation were classified with 85-100% accuracy except the kernels damaged by *T. castaneum* (73.3% accuracy). To further improve the classification accuracy, a two-way classification model was developed by combining the histogram features and the statistical features of the first PC wavelengths (1101.69 and 1305.05 nm) (Table 5.8). Since the wavelength (1447.46 nm) selected by the second PC loadings did not have any significance in the previous model, it was not included in this model. The developed two-way model correctly classified 85-100% of wheat kernels damaged by *T. castaneum*. Healthy wheat kernels and kernels damaged by *S. oryzae*, *R. dominica*, and *C. ferrugineus* were correctly identified with 91.7-100% accuracy using

**Table 5.3 Binary classification (20% test set) using histogram features (1101.69 nm wavelength) for healthy wheat kernels and those damaged by various insect species**

Sample (n=300)	Classification by discriminant analysis (%)			
	Linear		Quadratic	
	Crease-down	Crease-up	Crease-down	Crease-up
Healthy	100.0	95.0	100.0	98.3
<i>Sitophilus oryzae</i>	100.0	100.0	100.0	100.0
Healthy	98.3	95.0	88.3	93.3
<i>Rhizopertha dominica</i>	95.0	100.0	96.67	100.0
Healthy	100.0	91.7	100.0	98.3
<i>Cryptolestes ferrugineus</i>	90.0	93.3	95.0	91.7
Healthy	98.3	96.7	95.0	95.0
<i>Tribolium castaneum</i>	96.7	93.3	100.0	98.3

**Table 5.4 Binary classification (20% test set) using statistical (1101.69 and 1305.05 nm wavelengths) and histogram features (1101.69 nm wavelength) for healthy wheat kernels and those damaged by various insect species**

Sample (n=300)	Classification by discriminant analysis (%)			
	Linear		Quadratic	
	Crease-down	Crease-up	Crease-down	Crease-up
Healthy	100.0	100.0	100.0	100.0
<i>Sitophilus oryzae</i>	98.3	100.0	100.0	100.0
Healthy	100.0	100.0	90.0	96.7
<i>Rhizopertha dominica</i>	98.3	100.0	100.0	100.0
Healthy	100.0	100.0	100.0	98.3
<i>Cryptolestes ferrugineus</i>	100.0	98.3	91.7	96.7
Healthy	100.0	100.0	93.3	100.0
<i>Tribolium castaneum</i>	100.0	88.3	100.0	100.0

**Table 5.5 Two-way classification (20% test set) using statistical features (1101.69 and 1305.05 nm wavelengths) for healthy wheat kernels and those damaged by various insect species**

Sample (n=300)	Classification by discriminant analysis (%)			
	Linear		Quadratic	
	Crease-down	Crease-up	Crease-down	Crease-up
Healthy	96.7	100.0	90.0	85.0
<i>Sitophilus oryzae</i>	96.7	100.0	100.0	100.0
<i>Rhyzopertha dominica</i>	91.7	86.7	91.7	100.0
<i>Cryptolestes ferrugineus</i>	96.7	91.7	98.3	98.3
<i>Tribolium castaneum</i>	63.3	96.7	61.7	100.0

**Table 5.6 Two-way classification (20% test set) using statistical features (1101.69, 1305.05, and 1447.46 nm wavelengths) for healthy wheat kernels and those damaged by various insect species**

Sample (n=300)	Classification by discriminant analysis (%)			
	Linear		Quadratic	
	Crease-down	Crease-up	Crease-down	Crease-up
Healthy	95.0	85.0	86.7	86.7
<i>Sitophilus oryzae</i>	96.7	98.3	100.0	100.0
<i>Rhyzopertha dominica</i>	88.3	85.0	96.7	100.0
<i>Cryptolestes ferrugineus</i>	98.3	88.3	98.3	98.3
<i>Tribolium castaneum</i>	65.0	88.3	86.7	100.0

**Table 5.7 Two-way classification (20% tests set) using histogram features (1101.69 nm wavelength) for healthy wheat kernels and those damaged by various insect species**

Sample (n=300)	Classification by discriminant analysis (%)			
	Linear		Quadratic	
	Crease-down	Crease-up	Crease-down	Crease-up
Healthy	100.0	91.7	91.7	91.7
<i>Sitophilus oryzae</i>	100.0	100.0	100.0	100.0
<i>Rhizopertha dominica</i>	93.3	100.0	95.0	100.0
<i>Cryptolestes ferrugineus</i>	85.0	90.0	90.0	95.0
<i>Tribolium castaneum</i>	93.3	73.3	100	90.0

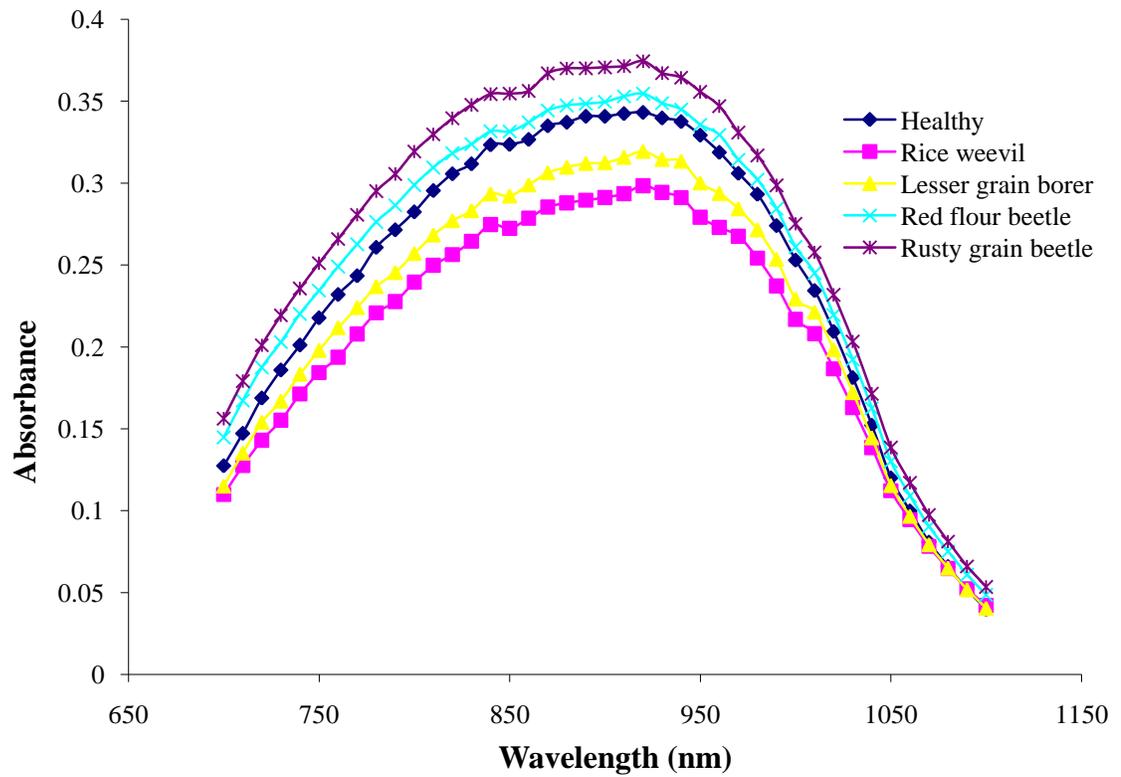
**Table 5.8 Two-way classification (20% tests set) using statistical (1101.69 and 1305.05 nm wavelengths) and histogram features (1101.69 nm wavelength) for healthy wheat kernels and those damaged by various insect species**

Sample (n=300)	Classification by discriminant analysis (%)			
	Linear		Quadratic	
	Crease-down	Crease-up	Crease-down	Crease-up
Healthy	100.0	96.7	91.7	93.3
<i>Sitophilus oryzae</i>	98.3	98.3	100.0	100.0
<i>Rhizopertha dominica</i>	93.3	100.0	98.3	100.0
<i>Cryptolestes ferrugineus</i>	95.0	93.3	100.0	100.0
<i>Tribolium castaneum</i>	88.3	85.0	98.3	100.0

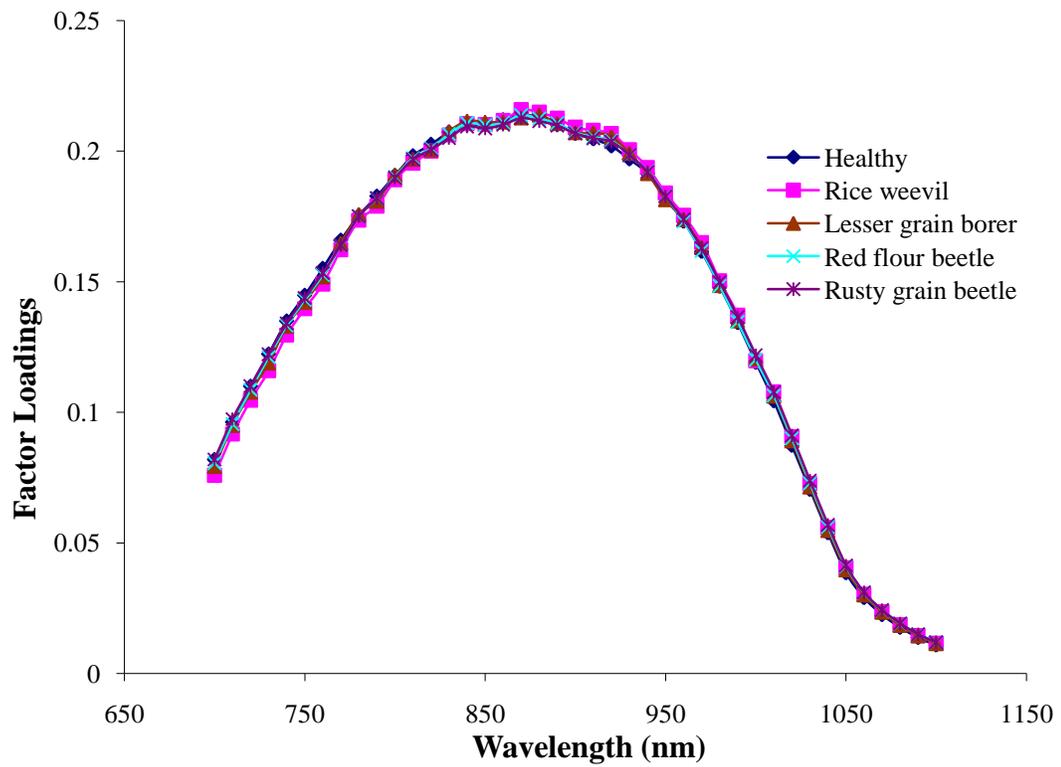
both LDA and QDA classifiers. Kernel orientation had no significant effect on the performance of the classification model which is another advantage over visual inspection. These two selected wavelengths (1101.69 and 1305.05 nm) can be used in real-time applications which would require less scanning time and only a few seconds for the classification of each sample.

### **5.1.2 Detection of insect-damaged kernels using SWNIR hyperspectral imaging**

The wavelength corresponding to the highest factor loading of first principal component which contained more than 98.7% variance was selected as the significant wavelength. Since kernel orientation did not significantly affect the classification and second PC loadings also did not improve the classification using LWNIR (section 5.1.1), these two parameters were not included in the analysis. The absorbance ( $\log(1/R)$ , where R is reflectance) spectra and first PC factor loadings of healthy and insect-damaged kernels are shown in Figs. 5.4 and 5.5. The 870 nm wavelength has the highest factor loading in the first PC and both absorbance spectra and factor loadings have similar shape. The shape of the spectra and loadings (bell-shaped) in SWNIR hyperspectral images showed wider and overlapping spectra compared to the spectra of LWNIR images which had many peaks and valleys in spectra and loadings. Due to the broad bell-shaped spectra in SWNIR imaging, only one wavelength was selected since addition of another wavelength would lead to a co-linearity problem with no further improvement in the classification. The wavelength 870 nm corresponds to the region of the third overtone of  $\text{CH}_3$  group. In a previous study, Perez-Mendoza et al. (2003) reported wavelength 890 nm as one of the important wavelengths from the analysis of beta coefficient using PLS for detection of



**Fig. 5.4 Short-wave near-infrared absorbance mean spectra of healthy and insect-damaged wheat kernels**



**Fig. 5.5 First principal component (PC) mean factor loadings of healthy and insect-damaged wheat kernels (short-wave near-infrared)**

insect fragments in wheat flour. Since binary classification models using LWNIR (section 5.1.1) gave high accuracy and indicated the discrimination capability of NIR imaging, the focus was to develop a realistic two-way classification model for sorting grain into healthy and damaged class by grouping insect-damaged kernels (all species) into one group and healthy kernels into another group. The two-way classification algorithm was developed using 13 features (6 statistical and 7 histogram features) from the 870 nm wavelength image. Only seven histogram features were used as the remaining histogram groups had zero values. The classification results of two-way classification are given in Table 5.9. Mahalanobis discriminant classifier gave the highest classification accuracy in detecting the insect-damaged kernels with 92.7-100% accuracy and 0.9-6.9 standard deviation. The kernels damaged by *T. castaneum* gave the lowest classification (92.7%) among the all insect species. Similar to LWNIR region, the kernels damaged by *T. castaneum* gave the lowest classification (Table 5.8). These two studies showed that kernels damaged by *T. castaneum* are the most difficult to detect by NIR hyperspectral imaging which might be due to the low damage caused to only the wheat germ area. The quadratic discriminant classifier gave the highest classification accuracy of 86% (22.0% std. dev) in classifying healthy kernels. The classification accuracy of LDA and Mahalanobis classifiers was lower in classifying healthy kernels. The LWNIR imaging system gave better classification in classifying healthy kernels.

**5.1.3 Detection of insect-damaged kernels using color imaging** Morphological, color, and textural image features were extracted from the digital color images. Among these features, color features are the simplest features in terms of computational complexity

**Table 5.9 Two-way classification of healthy and insect-damaged wheat kernels by four insect species using SWNIR hyperspectral image features**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	79.3	26.0	86.0	22.0	74.3	31.5
<i>S. oryzae</i>	97.3	3.0	99.3	1.5	100.0	0.0
<i>R. dominica</i>	97.7	3.7	91.3	5.8	98.7	1.8
<i>C. ferrugineus</i>	95.7	3.0	92.7	9.1	99.3	0.9
<i>T. castaneum</i>	90.3	6.9	83.3	8.4	92.7	6.9

and have also demonstrated high discrimination capability in many machine vision applications. So an attempt was made to develop statistical classification models using only top color features. Top 10 color features ranked using STEPDISC procedure are given in Table 5.10. These color features were given as input to develop two-way LDA, QDA, and Mahalanobis discriminant classifiers. In primary investigation, top 5 features gave lower classification and top 20 features did not improve the classification and in some cases resulted in non-positive covariance matrices and statistical classification failed. Therefore, two-way classification models were developed using the top 10 features. The classification results of various discriminant classifiers are given in Table 5.11. Linear discriminant classifier classified 88.0-99.0% insect-damaged kernels (0.9-10.1 std dev) followed by QDA and Mahalanobis classifiers which detected 83.7-98.3 (2.2-11.1 std dev) and 66.-94.7% (3.2-15.5 std dev) insect damaged-kernels, respectively. Kernels damaged by *C. ferrugineus* gave the lowest classification. The detection rate of healthy kernels was lower in all the three classifiers compared to damaged kernels. Linear

**Table 5.10 Top 10 features selected from 123 color features of healthy and insect-damaged wheat kernels**

<b>Rank</b>	<b>Feature</b>
1	Red moment 2
2	Red moment 3
3	Green moment 4
4	Red moment 1
5	Hue range
6	Green moment 1
7	Hue mean
8	Green mean
9	Green moment 3
10	Saturation range

**Table 5.11 Two-way classification of healthy and insect-damaged wheat kernels using top 10 color features extracted from color images**

<b>Sample</b>	<b>Classification by statistical discriminant classifiers (%)</b>					
	<b>Linear</b>		<b>Quadratic</b>		<b>Mahalanobis</b>	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	80.0	28.7	76.0	28.6	81.7	29.1
<i>S. oryzae</i>	99.0	0.9	98.3	2.04	94.7	3.2
<i>R. dominica</i>	97.0	2.2	98.0	2.2	89.7	8.2
<i>C. ferrugineus</i>	88.0	10.1	83.7	11.1	66.3	15.5
<i>T. castaneum</i>	91.0	6.2	92.3	5.1	86.3	9.4

**Table 5.12 Top 10 features selected from 230 morphological, color, and textural features of healthy and insect-damaged wheat kernels**

Rank	Feature	Feature
1	Red moment 3	Color
2	Red moment 2	Color
3	Red histogram range 5	Color
4	Blue moment 3	Color
5	Green moment 2	Color
6	Gray level run length matrix long run for green band	Textural
7	Gray level co-occurrence matrix uniformity for blue band	Textural
8	Saturation mean	Color
9	Blue moment 2	Color
10	Red histogram range 8	Color

**Table 5.13 Two-way classification of healthy and insect-damaged kernels using top 10 features from 230 morphological, color, and textural features extracted from color images**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	73.3	37.6	78.7	34.5	84.0	29.4
<i>S. oryzae</i>	99.3	0.9	98.3	2.4	94.0	4.2
<i>R. dominica</i>	99.3	1.5	96.3	3.8	85.7	10.4
<i>C. ferrugineus</i>	98.3	2.4	90.0	8.6	69.0	18.2
<i>T. castaneum</i>	93.7	5.8	85.7	8.8	75.3	14.0

discriminant classifier gave the highest classification accuracy of 80.0% (28.7 std dev) in discriminating healthy kernels from the insect-damaged kernels. To improve the classification, the top 10 features (Table 5.12) selected from combined 230 features (morphological, color, and textural) by SAS STEPDISC procedure were used to develop statistical discriminant classification models (Table 5.13). Linear discriminant classifier gave the highest classification accuracy and classified 93.7-99.3% insect-damaged kernels (5.8-0.9 std dev). Both QDA and Mahalanobis classifier gave lower classification compared to the LDA. The discrimination rate of healthy kernels was lower compared to the insect-damaged kernels. Mahalanobis classifier gave the highest accuracy of 84.0% (29.4 std dev) in discriminating healthy kernels from insect-damaged kernels. Though, the color machine vision system gave very high classification accuracy in classifying insect-damaged kernels, the identification of healthy kernels gave very high false positive error and nearly 20-30% kernels were misclassified as damaged kernels.

**5.1.4 Detection of insect-damaged kernels using color and SWNIR imaging** Since the use of only SWNIR images or color images resulted in high false positive error despite the highest classification accuracy of damaged kernels, an attempt was made to overcome the false positive error in classifying the healthy kernels. The features from color images (both color and combined features) were combined with SWNIR image features. Combining the features was possible because samples were labelled and scanned simultaneously during image acquisition using both the color and SWNIR imaging systems. First, the classification algorithms were developed by combining the top 9 color features (10<sup>th</sup> feature removed due to non-covariance problem) with the shortwave NIR image features (Table 5.14). Linear discriminant classifier gave the highest classification

accuracy and classified more than 96.0% damaged kernels (0.9-2.6 std dev) and 89.0% healthy kernels (13.1 std dev), followed by QDA and Mahalanobis classifiers. Both QDA and Mahalanobis discriminant classifiers also identified 93.0-100% (0-7.4 std dev) and 97-100.0% (0-4.5 std dev) insect-damaged kernels, respectively, but gave false positive error of more than 10.0% in classifying healthy kernels. To investigate further improvement in performance and reduce the false positive error, classification algorithms were developed using the top 10 features from 230 combined features (morphological, color, and textural) and SWNIR image features (Table 5.15). All the three classifiers gave very high accuracy in classifying the insect-damaged kernels. More than 97.0% of kernels damaged by *S. oryzae*, *R. dominica*, and *C. ferrugineus* were correctly classified with only 0-1.9 standard deviation. The classification accuracy of kernels damaged by *T. castaneum* was slightly lower (91.0-95.3%). These studies showed that the kernels damaged by *T. castaneum* are the most difficult to identify using imaging systems in visible and NIR region of electromagnetic spectrum. The QDA gave highest classification of 96.3% (4.5 std dev) in discriminating healthy kernels from insect-damaged kernels. Overall accuracy of QDA was more than 91.0% in classifying healthy and insect-damaged kernels.

**5.1.5 Classification of insect-damaged kernels using BPNN** To assess if the classification accuracy can be improved and false positive error can be reduced using BPNN with SWNIR and color imaging systems, classification algorithms were also developed for three-layer BPNN classifier (Table 5.16). The BPNN classifier using SWNIR image features as input classified 95.0-99.7% (0.8-4.6 std dev) damaged kernels,

**Table 5.14 Two-way classification of healthy and insect-damaged wheat kernels using combined top 9 features from 123 color features and SWNIR image features**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	89.3	13.1	86.3	15.5	83.0	17.7
<i>S. oryzae</i>	99.3	0.9	100.0	0.0	100.0	0.0
<i>R. dominica</i>	99.3	1.5	99.3	0.9	99.3	0.9
<i>C. ferrugineus</i>	99.3	0.9	98.0	2.2	99.7	0.8
<i>T. castaneum</i>	96.7	2.6	93.0	7.4	97.0	4.5

**Table 5.15 Two-way classification of healthy and insect-damaged kernels using combined top 10 features from 230 morphological, color, and textural features and SWNIR image features**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	81.3	26.8	96.3	4.5	92.7	9.3
<i>S. oryzae</i>	100.0	0.0	100.0	0.0	100.0	0.0
<i>R. dominica</i>	99.7	0.8	97.7	1.9	98.7	1.4
<i>C. ferrugineus</i>	99.7	0.8	99.3	0.9	100.0	0.0
<i>T. castaneum</i>	95.3	7.0	91.0	9.6	92.7	8.2

**Table 5.16 Two-way classification of healthy and insect-damaged wheat kernels by back-propagation neural network (BPNN) classifier**

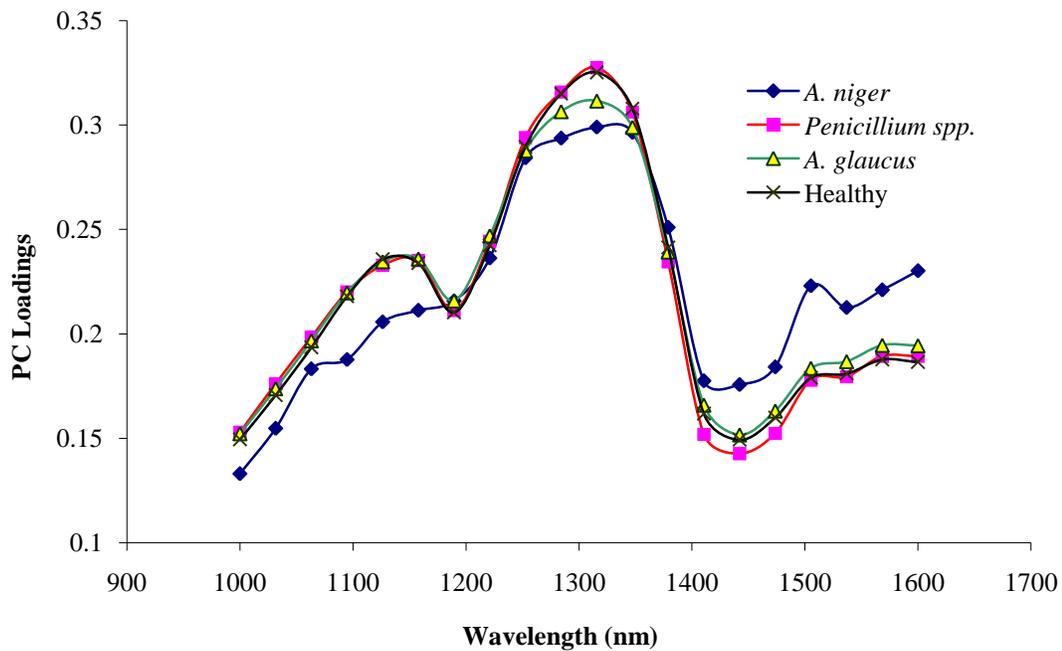
Sample	Classification by using various features					
	NIR Image		Color Image		Combined	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	72.3	31.5	75.0	34.8	87.0	12.8
<i>S. oryzae</i>	98.3	2.4	99.0	1.5	99.7	0.8
<i>R. dominica</i>	99.7	0.8	98.0	2.7	98.3	3.7
<i>C. ferrugineus</i>	98.3	1.9	97.7	2.5	99.0	0.9
<i>T. castaneum</i>	95.0	4.6	97.0	2.2	95.7	5.4

however, healthy kernels were highly misclassified (72% accuracy). Similarly, BPNN classifier using the top ten features from 230 morphological, color, and textural features also misclassified nearly 25.0% healthy kernels but gave high insect damage detection rates of 97.0-99.0% (1.5-2.7 std dev). The BPNN classification model was also developed by combining the NIR image features and the top 10 features from color images. The BPNN model using combined features classified 95.7-99.7% insect-damaged kernels (0.8-5.4 std dev) and 87.0% healthy kernels (12.8 std dev).

The BPNN models did not show any improvement over statistical discriminant classifiers. The statistical classifiers also have certain advantage such as easy and fast training compared to neural network classifiers. Therefore, classification results of only statistical classifier are discussed in later sections.

## 5.2 Fungal Damage Detection

**5.2.1 Fungal detection using LWNIR hyperspectral imaging** Out of 20 wavelength bands in the range of 1000-1600 nm; wavelengths of 1284.2, 1315.8, and 1347.4 nm had the highest factor loadings corresponding to the first principal component (Fig. 5.6) (Singh et al., 2007). These three wavelengths were considered as the most significant wavelengths and were used for further classification by k-means classification and discriminant analysis. Wang et al. (2003) found 1330 nm as one of the significant wavelengths in their study to classify fungal-damaged soybean seeds, and related it to the CH bond representing fiber and starch content of fungal-damaged soybean seeds. I did not directly apply the k-means classification and discriminant analysis methods using reflectance values from all the pixels in a kernel due to variation in kernel size and thus total number of pixels. Both the k-means and discriminant classification methods require matrices with the same number of elements in each row (each row corresponds to a sample image). So the mean, maximum and minimum reflectance intensities from the image slices corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths were used in classification. In this analysis twenty wavebands were used, which took nearly 1.5 min for scanning, pre-processing, labeling, reshaping and applying PCA to all five kernels of an image. However, future analysis would be very fast as just three scans at significant wavelengths will be required for classification. First, an attempt was made to see if the kernels infected from an individual fungal species could be separated from healthy kernels. A pair-wise k-means classifier was tested using 600 kernels (300 from each of three fungi and 300 healthy kernels) for each analysis of *Penicillium*-infected versus



**Fig. 5.6 First principal component (PC) loadings of healthy and fungal infected wheat kernels (long-wave near-infrared)**

healthy, *A. glaucus*-infected versus healthy, and *A. niger*-infected versus healthy kernels (Table 5.17). The classification accuracy varied from 61% to 82.67 % for infected kernels and 80 to 84% for healthy kernels. Pair-wise discriminant classifiers (linear, quadratic and Mahalanobis) with an independent sample set of 300 infected and 300 healthy kernels (60 test set and 240 training set for each) were also used (Table 5.18) to separate kernels infected with each fungal species from healthy kernels. All the kernels infected with *Penicillium* were correctly classified against the healthy kernels in the test set, of which more than 95% healthy kernels were correctly classified. The maximum classification accuracy of kernels infected with *A. niger* was 95% against the healthy kernels (more than 98.3 % accuracy) in both linear and Mahalanobis discriminant analyses. Classification accuracy of kernels infected with *A. glaucus* group was lower (91.7%) as compared to kernels infected with *Penicillium* and *A. niger* and approximately 5-7% healthy kernels were misclassified as kernels infected with *A. glaucus*. Overall, discriminant analysis gave better results as compared to k-means classification. So the k-means classifier was not used in further classification analysis.

In another two-way discriminant classification analysis, kernels were classified as infected or healthy with sample set of 1200 kernels (60 kernels as test set and 240 kernels as training set from each of fungal-infected and healthy kernels) (Table 5.19). In all the samples sets independent samples were used for training and test (true-validation). All the tables contain classification results from test set data only. All the kernels infected with *Penicillium* spp. were correctly classified as infected kernels by all three discriminant classifiers (linear, quadratic, and Mahalanobis). Linear discriminant classifier gave the highest classification accuracy and on average about 97.8% infected

**Table 5.17 Pair-wise discrimination of healthy and fungal-infected wheat kernels by k-means classifier using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths**

<b>Fungi</b>	<b>Sample Size</b>	<b>Classification accuracies (%)</b>	
<i>Penicillium</i>	(no. of kernels) 600	<i>Penicillium</i> 82.7	Healthy 84
<i>A. niger</i>	600	<i>A. niger</i> 61	Healthy 81.3
<i>A. glaucus</i>	600	<i>A. glaucus</i> 69.3	Healthy 80

**Table 5.18 Pair-wise discrimination of healthy and fungal-infected wheat kernels by discriminant analysis using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths**

<b>Fungi</b>	<b>Classification accuracies of various discriminant classifiers (%)</b>					
	<b>Linear</b>		<b>Quadratic</b>		<b>Mahalanobis</b>	
<i>Penicillium</i>	<i>Penicillium</i> 100	Healthy 95	<i>Penicillium</i> 100	Healthy 98.3	<i>Penicillium</i> 100	Healthy 95
<i>A. niger</i>	<i>A. niger</i> 95	Healthy 100	<i>A. niger</i> 91.6	Healthy 100	<i>A. niger</i> 95	Healthy 98.3
<i>A. glaucus</i>	<i>A. glaucus</i> 88.3	Healthy 93.3	<i>A. glaucus</i> 91.7	Healthy 95	<i>A. glaucus</i> 91.7	Healthy 95

kernels were correctly classified. However, about 5% healthy kernels were misclassified as infected. Quadratic and Mahalanobis discriminant models, on average, correctly classified 93.3% and 96.1% infected kernels, respectively. The performance of these two classifiers was slightly lower than that of linear discriminant classifier. Linear discriminant classifier uses the same covariance (pooled covariance) for each class whereas quadratic discriminant classifier calculates a covariance matrix for each class separately. Better performance by linear discriminant classifier could be attributed to the equal dispersion of reflectance intensities in all the fungal classes. Pearson and Wicklow (2006) did a comparative study to detect fungal-damaged corn kernels using physical properties, NIR transmittance images, color transmittance images, color reflectance images, X-ray images and NIR spectra and identified fungal-infected corn kernels with a maximum accuracy of 91% using NIR spectra (75-91% overall classification range) and detected healthy kernels with classification accuracy of 90 to 100%. Two-way linear discriminant classifier used in this study gave better results compared to these methods. The classification accuracy of the above discussed classifiers was also tested using mixed training set containing 20% infected and 80% healthy kernels (total 60 kernels) to detect the infected kernels from healthy kernels (Table 5.20). Both the linear and quadratic classifiers classified the kernels from the mixed samples of healthy and infected kernels with an average classification accuracy of 98.3 and 98.9%, respectively. The Mahalanobis classifier correctly classified more 95.6% kernels in the test set. Misclassification occurred as some of the healthy kernels were classified as infected kernels. The classification algorithms discussed above were only able to identify whether the kernels were healthy or infected. An attempt was also made to differentiate between

**Table 5.19 Two-way discriminant classification analysis to separate fungal infected kernels from healthy kernels using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths**

Actual class	Classification accuracies of various discriminant classifiers (%)		
	Linear	Quadratic	Mahalanobis
Healthy	<b>95</b>	<b>96.7</b>	<b>93.3</b>
<i>Penicillium</i>	100	100	100
<i>A. niger</i>	96.7	93.3	96.7
<i>A. glaucus</i>	96.7	86.7	91.7
<b>Average</b>	<b>97.8</b>	<b>93.3</b>	<b>96.1</b>

**Table 5.20 Two-way classification analysis using test set of 60 mixed kernels (48 healthy and 12 infected kernels) and highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths**

Actual class	Classification by discriminant classifier (%)		
	Linear	Quadratic	Mahalanobis
Healthy + <i>Penicillium</i>	98.3	98.3	95
Healthy + <i>A. niger</i>	98.3	98.3	95
Healthy + <i>A. glaucus</i>	98.3	100	96.7
<b>Average</b>	<b>98.3</b>	<b>98.9</b>	<b>95.6</b>

two kinds of infections, i.e., *Penicillium* versus *A. niger*, *Penicillium* versus *A. glaucus*, and *A. niger* versus *A. glaucus* (Table 5.21). The results showed that kernels infected with *Penicillium* could be correctly differentiated from *A. niger* and *A. glaucus* group infected kernels with maximum classification accuracy of 100% and 96.7%, respectively. However, kernels infected with *A. glaucus* were misclassified as *Penicillium* and gave very poor classification results (less than 75%). Kernels infected with *A. niger* and *A. glaucus* were misclassified with each other. From all the classification models discussed above it was observed that the infected kernels can be well differentiated from the healthy kernels. However, these classifiers were unable to differentiate between *A. niger* and *A. glaucus* infections. The misclassification occurred because there might not be significant difference among the pixel intensities of various types of fungal infected kernels. However, there was a significant difference between pixel intensities of healthy and infected kernels to separate them. The fungal infected kernels may have lower lipid and starch content compared with healthy kernels (Madhyastha et al., 1993; Daftary and Pomeranz, 1965). Due to differences in these chemical compositions, fungal-infected kernels were well separated from healthy kernels but there might not be significant compositional difference in the kernels infected with the species of same genus (*A. niger* and *A. glaucus*) to be separated by hyperspectral image analysis. However, there is color difference between these two species of the same genera, which could be differentiated by using visible color images. *A. niger* has black spores and *A. glaucus* has blue-gray spores. Apart from two-way classification, four-class classification models were also developed to classify the kernels into healthy and different kinds of infection using the above discriminant classifiers. All the three discriminant classifiers gave mixed

classification results. The linear discriminant classifier correctly classified healthy kernels and kernels infected with *Penicillium* with classification accuracy of more than 95% and 91.7%, respectively (Table 5.22). Approximately 8% of kernels infected with *Penicillium* were misclassified as *A. glaucus* infected kernels. Some of the healthy kernels (approximately 5%) were misclassified as infected kernels. Kernels infected with *A. niger* and *A. glaucus* gave poor classification results with maximum classification accuracy of 60% and 41.7%, respectively. Quadratic discriminant classifier classified healthy kernels and kernels infected with *Penicillium*, *A. niger*, and *A. glaucus* with accuracies of 95%, 78.3%, 68.3% and 58.3%, respectively (Table 5.23). The Mahalanobis discriminant classifier correctly classified 93.3% healthy kernels, 96.7% *Penicillium* infected kernels and 75% kernels infected with *A. niger* (Table 5.24). However, this classifier failed to detect *A. glaucus* infected kernels (30% accuracy) and 53.3% of them were misclassified as kernels infected with *Penicillium*.

**5.2.2 Detection of fungal-infected kernels using SWNIR hyperspectral imaging** The significant wavelength corresponding to the highest factor loading of first principal component (more than 98.3% variation) was selected as the significant. Since kernel orientation did not significantly affect the classification and second PC loadings did not improve the classification (section 5.1.1), these two parameters were not included in this analysis. The absorbance spectra and PC loadings of healthy and fungal-infected wheat kernels are shown in Figs. 5.7 and 5.8, respectively. The 870 nm wavelength corresponding to the highest factor loading of first PC was selected as the significant

**Table 5.21** Pair-wise discrimination between kernels infected with various types of fungi by discriminant classifiers using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths

Classifier	Classification accuracies of samples (%)					
	<i>Penicillium</i> vs <i>A. niger</i>		<i>Penicillium</i> vs <i>A. glaucus</i>		<i>A. niger</i> vs <i>A. glaucus</i>	
	<i>Penicillium</i>	<i>A. niger</i>	<i>Penicillium</i>	<i>A. glaucus</i>	<i>A. niger</i>	<i>A. glaucus</i>
Linear	98.3	90.0	96.7	45.0	66.7	93.3
Quadratic	100	85.0	78.3	75.0	68.3	86.7
Mahalanobis	98.3	80.0	96.7	43.3	86.7	80.0

**Table 5.22.** Four-class classification (Linear discriminant analysis) using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths

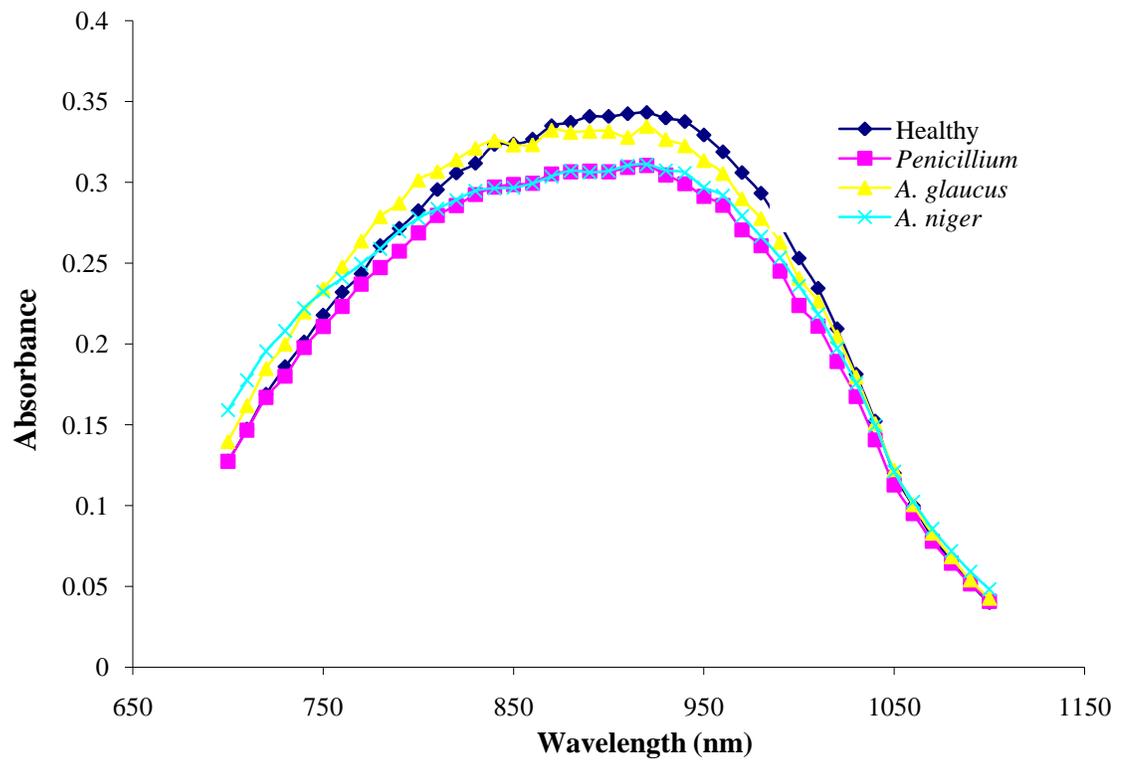
Actual class	Classification accuracies of samples (%)			
	Healthy	<i>Penicillium</i>	<i>A. niger</i>	<i>A. glaucus</i>
Healthy	<b>95</b>	3.3	0	1.7
<i>Penicillium</i>	0	<b>91.7</b>	0	8.3
<i>A. niger</i>	1.7	8.3	<b>60</b>	30
<i>A. glaucus</i>	3.3	50	5	<b>41.7</b>

**Table 5.23 Four-class classification (Quadratic discriminant analysis) using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths**

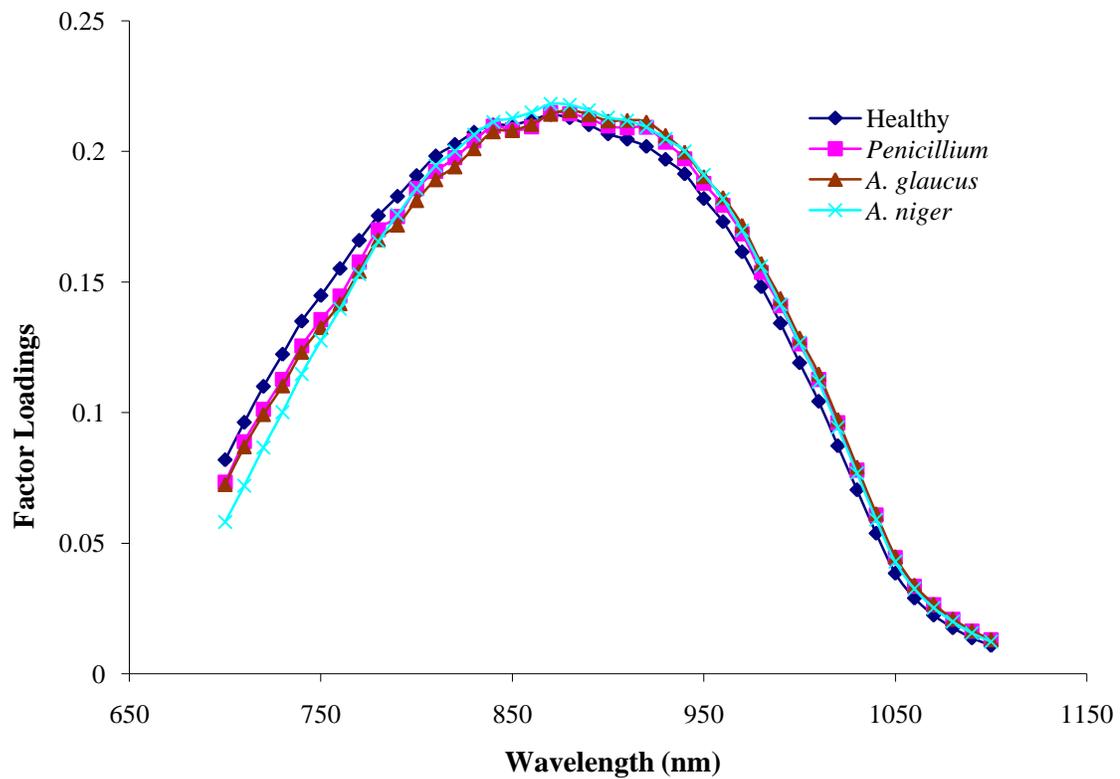
Actual class	Classification accuracies of samples (%)			
	Healthy	<i>Penicillium</i>	<i>A. niger</i>	<i>A. glaucus</i>
Healthy	<b>95</b>	0	0	5
<i>Penicillium</i>	0	<b>78.3</b>	0	21.7
<i>A. niger</i>	3.3	5	<b>68.3</b>	23.3
<i>A. glaucus</i>	8.3	20	13.3	<b>58.3</b>

**Table 5.24 Four-class classification (Mahalanobis discriminant analysis) using highest, lowest and mean reflectance values corresponding to 1284.2, 1315.8, and 1347.4 nm wavelengths**

Actual class	Classification accuracies of samples (%)			
	Healthy	<i>Penicillium</i>	<i>A. niger</i>	<i>A. glaucus</i>
Healthy	<b>93.3</b>	3.3	0	3.3
<i>Penicillium</i>	0	<b>96.7</b>	0	3.3
<i>A. niger</i>	3.3	15	<b>75</b>	6.7
<i>A. glaucus</i>	6.7	53.3	10	<b>30</b>



**Fig. 5.7 Short-wave near-infrared absorbance mean spectra of healthy and fungal-infected wheat kernels**



**Fig. 5.8 First principal component (PC) mean factor loadings of healthy and fungal-infected wheat kernels (short-wave near-infrared)**

wavelength. Only one wavelength in this region was selected as addition of another wavelength would create co-linearity problem. The wavelength 870 nm corresponds to the region of third overtone of CH<sub>3</sub> group. In a previous study, Perez-Mendoza et al. (2003) reported wavelength 890 nm as one of the important wavelengths from the analysis of beta coefficient using PLS for detection of insect fragments in wheat flour. The two-way classification algorithms were developed using 13 features (6 statistical and 7 histogram features) from the 870 nm waveband image to discriminate fungal-infected wheat kernels from healthy kernels. Only seven histogram features were used as remaining histogram groups had zero values. Four-class models were also developed to classify each fungal species and healthy kernels. However, four-class models failed to correctly classify individual fungal-infected species and resulted in high misclassification in four-way classification so the results are not discussed but are given in appendix (Table A4-A6). The two-way classification results of LDA, QDA, and Mahalanobis classifiers are given in Table 5.25. All the three classifiers gave very high accuracy in classifying fungal-infected kernels. Both LDA and QDA classified 95.7-98.0% (2.2-5.5 std dev) and 96.0-98.3% (1.2-4.4 std dev) fungal-infected kernels, respectively. Mahalanobis classifier classified 94.0-96.7% (2.4-6.3 std dev) fungal-infected kernels. Healthy wheat kernels were correctly identified by LDA, QDA, and Mahalanobis classifiers with classification accuracies of 87.7% (7.9 std dev), 91.7% (5.0 std dev), and 93.7% (5.2 std dev), respectively. Classification performance of SWNIR imaging is similar to the LWNIR imaging classification given in the previous section (section 5.2.1). However, the cost of SWNIR imaging system (CCD camera) would be lower than the LWNIR imaging system (InGaAs Camera) due to camera cost. This would be one of the

major criteria in selection of an appropriate imaging system. Most of the pushbroom type hyperspectral imaging systems (line-scan) use CCD cameras operating in the range of 700-1100 nm wavelength. Color cameras are very inexpensive compared to SWNIR and LWNIR imaging systems. The color imaging can also potentially discriminate fungal-infected and healthy wheat kernels, so the potential of color digital imaging was also investigated to detect fungal-infected wheat kernels.

**5.2.3 Detection of fungal-infected kernels using color imaging** A total of 230 features (morphological, color, and textural features) from the images of fungal-infected and healthy wheat kernels were extracted. Among these 230 extracted features; color features were considered the simplest features in terms of computational complexity and had also demonstrated high discrimination capability in many machine vision applications. So, statistical classification models for fungal detection were first developed using only the top 10 color features. The top 10 color features ranked using STEPDISC procedure are given in Table 5.26. These color features were given as input to develop two-way LDA, QDA, and Mahalanobis discriminant classifiers. In primary investigation, the top 5 features gave lower classification and more than 8 features in some cases resulted in non-positive covariance matrices due to which statistical classification failed. Therefore, two-way classification models were developed using the top 8 features from 123 color features. The two-way classification results of classifying fungal-infected kernels using these features are given in Table 5.27. All the three classifiers gave high accuracies in detecting fungal-infected wheat kernels. All the kernels infected by *A. niger* were correctly identified by three classifiers. Kernels infected by *Penicillium* spp. and *A. glaucus* were also correctly identified with 99.3 (0.9 std dev) and 98.0% (2.7 std dev)

**Table 5.25 Two-way classification of healthy and fungal-infected wheat kernels using SWNIR hyperspectral image features**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	88.7	7.9	91.7	5.0	93.7	5.2
<i>Penicillium</i>	98.0	2.2	98.3	1.2	96.7	2.4
<i>A. niger</i>	95.7	5.5	96.0	4.4	94.0	6.3
<i>A. glaucus</i>	96.7	2.4	96.7	2.0	94.7	3.2

**Table 5.26 Top 10 features selected from 123 color features of healthy and fungal-infected wheat kernel**

Rank	Feature
1	Red moment 2
2	Green moment 2
3	Hue mean
4	Blue range
5	Red range
6	Saturation mean
7	Blue variance
8	Red variance
9	Red moment 3
10	Green moment 3

**Table 5.27 Two-way classification of healthy and fungal-infected wheat kernels using top 8 color features from 123 color features**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	92.7	7.2	90.0	11.37	85.7	13.6
<i>Penicillium</i>	98.3	1.8	98.7	1.8	99.3	0.9
<i>A. niger</i>	100.0	0.0	100.0	0.0	100.0	0.0
<i>A. glaucus</i>	94.3	3.3	89.3	9.1	98.0	2.7

**Table 5.28 Top 10 features selected from 230 morphological, color, and textural features of healthy and fungal-infected wheat kernels**

Rank	Feature	Feature type
1	Gray level run length matrix long run for green band	Textural
2	Gray level co-occurrence matrix uniformity for blue band	Textural
3	Red histogram range 12	Color
4	Gray level co-occurrence matrix uniformity for red band	Texture
5	Gray level co-occurrence matrix uniformity for green band	Texture
6	Blue variance	Color
7	Red variance	Color
8	Saturation range	Color
9	Peripheral Fourier descriptor 20	Morphological
10	Peripheral Fourier descriptor 17	Morphological

accuracies, respectively, by Mahalanobis classifier. Healthy kernels were correctly classified with 92.7 (7.2 std dev), 90.0 (11.4 std dev), and 85.7 % (13.6 std dev) accuracy by LDA, QDA, and Mahalanobis classifiers, respectively.

Two-way classification algorithms were also developed using the top 10 selected features from 230 morphological, color, and textural features. The top 10 features selected using STEPDISC procedure, are given in Table 5.28. Top 10 feature set consisted of color, textural, and morphological features. The classification results of algorithms developed using these features as input are given in Table 5.29. These results are similar to the classification using color features (Table 5.27) with slight variation. Kernels infected by *Penicillium* and *A. niger* were correctly identified with more than 97.0% accuracy (0.8-4.4 std dev) by all the three classifiers. The classification accuracy of identifying *A. glaucus*-infected kernels by Mahalanobis classifier was reduced to 94.0%. The performance of classifying healthy kernels by LDA improved from 92.7% (7.2 std dev) to 95.7% (7.3 std dev). This improvement in classification of healthy kernels indicates that inclusion of textural and morphological features reduces the false positive error, whereas color based information is more significant for identification of fungal-infected kernels.

#### **5.2.4 Detection of fungal-infected kernels using color and SWNIR imaging**

Classification algorithms were also developed by combining the image features from color and SWNIR images. The First top 8 color features selected from 123 color features were combined with NIR image features and two-way classification algorithms were developed by statistical classifiers (Table 5.30). The classification accuracy of all fungal-infected kernels highly improved and more than 99.0% (0-.09 std dev) infected kernels

were correctly classified by LDA, QDA, and Mahalanobis classifiers. The classification accuracy of healthy kernels also improved and 96.0-97.7% (2.5-4.4 std dev) kernels were correctly identified by discriminant classifiers. The classification results of two-way classification using the top 10 features from 230 morphological, color, and textural features combined with SWNIR image features are given in Table 5.31. The LDA achieved near maximum classification accuracy with correct classification of 99.0-100% (0-0.9 std dev) in classifying healthy wheat kernels and kernels infected by all three fungal species. The QDA and Mahalanobis classifiers also gave very high accuracy of 98.7-99.0% (0.9-1.4 std dev) in indentifying fungal-infected kernels. Healthy kernels were identified with classification accuracies of 97.0 (3.0 std dev) and 97.7% (3.3 std dev) by QDA and Mahalanobis classifier, respectively. The results demonstrated the capability of improving the classification performance by combining features from color images and SWNIR images against using individually. The results from SWNIR images showed very high false positive error which can be eliminated by combining the SWNIR image features with color image features. These results were also comparable with LWNIR imaging classification and showed slight improvement in classification and less than 1% false positive error.

### **5.3 Detection of Sprout Damage**

**5.3.1 Detection of sprouted and midge-damaged kernels using LWNIR hyperspectral imaging** Two wavelength regions with dominating peaks near 1100 and 1300 nm had the highest first PC factor loadings (Singh et al., 2009b). Loadings of the principal components explain the correlation between variables and principal component. Variables with highest loadings value have highest contribution in that component and

**Table 5.29. Two-way classification of healthy and fungal-infected kernels using top 10 features from 230 morphological, color, and textural features**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	95.7	7.3	90.0	13.3	86.7	17.8
<i>Penicillium</i>	97.7	1.9	99.3	1.5	99.7	0.8
<i>A. Niger</i>	98.3	2.9	97.3	4.4	98.7	2.2
<i>A. Glaucus</i>	94.0	5.1	93.0	5.6	94.3	6.3

**Table 5.30 Two-way classification of healthy and fungal-infected kernels using combined top 8 color features from 123 color features and SWNIR hyperspectral image features**

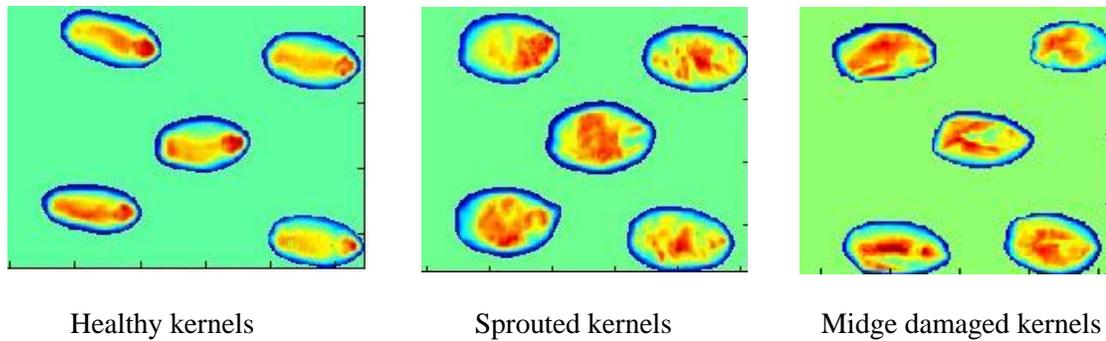
Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	96.0	4.4	97.0	2.7	97.7	2.5
<i>Penicillium</i>	100.0	0.0	99.0	0.9	99.0	0.9
<i>A. Niger</i>	99.3	0.9	99.3	0.9	99.3	0.9
<i>A. Glaucus</i>	99.7	0.8	99.0	0.9	99.0	0.9

**Table 5.31 Two-way classification of healthy and fungal-infected kernels using combined top 10 features from 230 morphological, color, and textural and SWNIR image features**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	99.0	0.9	97.0	3.0	97.3	3.3
<i>Penicillium</i>	100.0	0.0	99.0	0.9	99.0	0.9
<i>A. Niger</i>	99.7	0.8	99.0	1.4	99.0	1.4
<i>A. Glaucus</i>	99.0	0.9	98.7	1.4	98.7	1.4

are used to interpret the significance of the variables. Analysis of loadings gives more reliable and accurate results than rank estimation method based on eigenvalues criteria (Sasic, 2001; Rossi and Warner, 1986). The first PC accounted for 95-98% variability of data in healthy, artificially sprouted, and midge-damaged kernels. The first PC score images (pseudo color image) showed the obvious differences in compositional distribution between the damaged and healthy kernels (Fig. 5.9). The PC analysis gives scores matrix of size  $k \times 60$  as one of the outputs in which each column consists of PC scores of each PC from first to the last (60<sup>th</sup>) component. The actual intensities were replaced by scores of first PC (column 1) in the labeled kernels to form the pseudo color images. The kernels whose score images are shown in Fig. 5.9 were imaged with germ facing right side and crease-down position. The healthy kernels have solid contours and the germ area is clearly visible with relatively higher and concentrated intensity; however, in sprouted and midge-damaged kernels the germ portion does not show that pattern. This could be due to the starch decomposition or embryo development in the germ area of the sprouted and midge-damaged kernels.

In the discrimination of artificially sprouted and healthy kernels two wavelengths (1101.7 and 1305.1 nm) corresponding to the highest factor loadings of the first PC were selected as the significant wavelengths. In the classification of midge-damaged samples, three wavelengths (1101.7, 1132.20, and 1305.1 nm) were selected as significant wavelengths. In midge-damaged samples, first two wavelengths (1101.7, 1132.20 nm) in the 1100 nm peak region were selected due to the small shift in highest factor loadings with respect to the wavelengths for different location samples (Fig. 5.10). The significance of selected wavelengths can also be observed from the reflectance spectra as



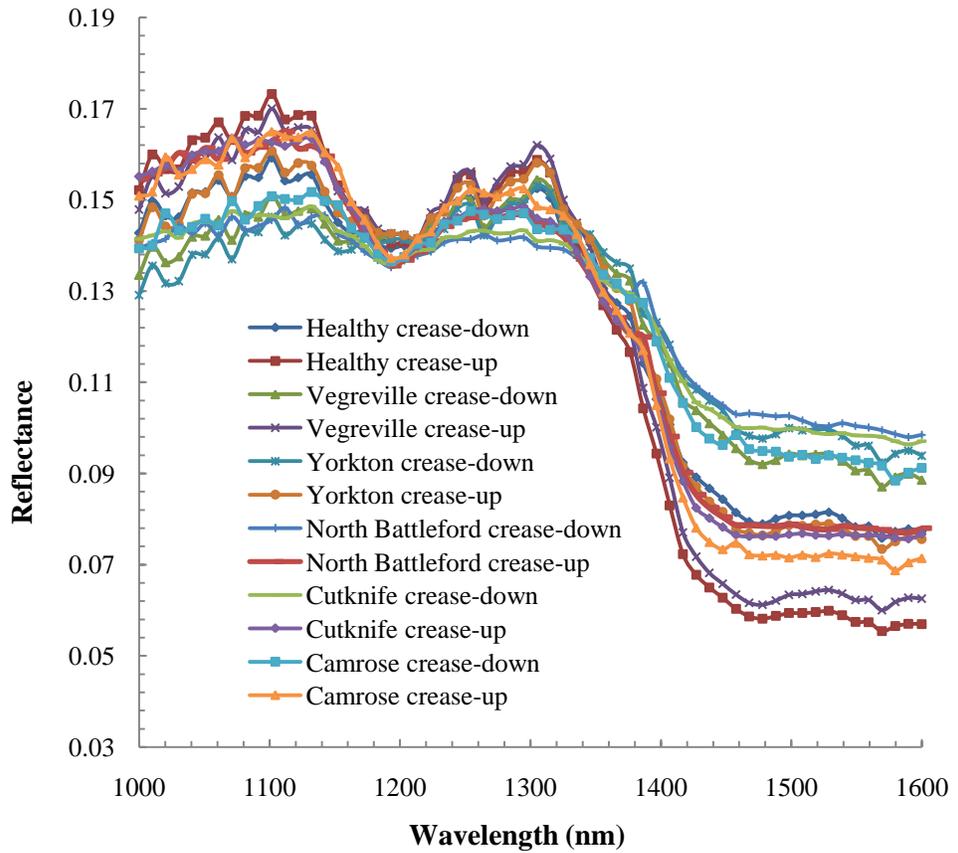
**Fig. 5.9 First principal component (PC) scores images of healthy, sprouted, and midge-damaged wheat kernels (long-wave near-infrared)**

these three wavelengths correspond to the main dominating peaks (Fig. 5.11). The wavelengths 1101.7 and 1132.2 nm correspond to the second overtone of CH and 1305.1 nm corresponds to the first overtone of CH combination bands. The significance of these wavelengths can be associated with absorption by starch molecules. The wavelength region examined in our study was also found significant by other researchers. Delwiche (1998) found the spectral region from 1100 to 1400 nm as the most significant region for protein content analysis of wheat and interpreted 1138 nm as the wavelength attributable to the protein absorption band. In another study, Delwiche and Hareland (2004) found the spectral region from 1130 to 1190 nm to be very stable for classifying normal and scab-damaged wheat kernels. Xie et al. (2004) related 1155 nm wavelength to the starch structure changes which correspond to the second overtone of the CH<sub>3</sub> bond. Barton and Burdick (1979) related the peak around the 1330 nm CH bond to the fiber and starch content.

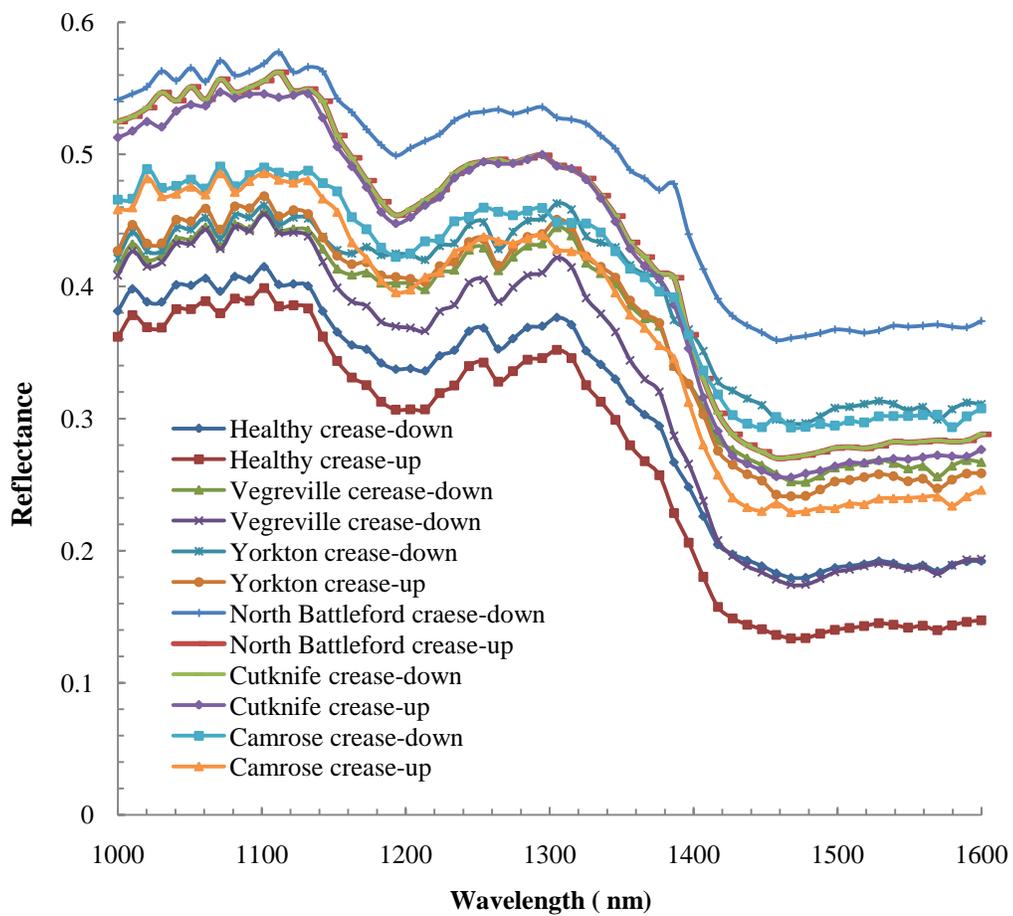
A binary classification model was developed to classify artificially sprouted wheat kernels and healthy wheat kernels (Table 5.32). Both LDA and QDA classifiers correctly classified healthy and sprouted kernels. Mahalanobis classifiers also classified

all the sprouted kernels and more than 98% healthy kernels. This high accuracy is expected as very low falling number value (high alpha amylase level) in the sprouted kernels was determined. In actual field conditions, the degree of sprout damage varies depending on weather conditions and level of midge-damage. Though, the midge-damage usually occurs on the pericarp, it does not happen all the time. The distribution of alpha-amylase within individual wheat kernels is heterogeneous. So the artificial sprout damage does not always simulate the actual sprout damage in the field.

To develop a robust classification model, I used the midge-damaged wheat samples from five different locations across western Canada. Due to the non-uniform nature of midge damage, the damaged samples were scanned both in crease-down and crease-up orientation. A two-way discriminant classification model for crease-down orientation was developed using healthy sample and midge-damaged samples from all five locations. Only one CWRS sample of healthy kernels was used as control sample because in Canada wheat of the same class from different growing locations is mixed in grain elevators and marketed as a composite class and is not segregated based on the growing regions. For the development of a real-time inspection system for grain in elevators, it is not practically significant to consider the growing location. The model was developed using an independent training set (combining 80% kernels from each sample) and tested by each of the sample (20% kernels remaining kernels from each sample). A similar classification model was also developed for crease-up orientations. The classification results of the crease-down orientation samples are given in Table 5.33. More than 95% of healthy kernels were correctly classified by LDA and QDA classifiers whereas Mahalanobis classifier correctly classified 85% of the healthy kernels. The LDA



**Fig. 5.10 PC loadings of first principal component (PC) of healthy and midge-damaged wheat kernels for different locations in western Canada (long-wave near-infrared)**



**Fig. 5.11 Long-wave near-infrared mean reflectance spectra of healthy and midge-damaged wheat kernels for different locations in western Canada**

**Table 5.32 Classification of sprout-damaged wheat kernels using 300 healthy and 300 sprouted kernels (80% training and 20% test sets) (LWNIR)**

Discriminant classifier	Classification accuracy (%)	
	Healthy	Sprouted
Linear	100.0	100.0
Quadratic	100.0	100.0
Mahalanobis	98.3	100.0

and QDA classifiers gave moderate to high classification accuracy (83.33-100%) in classification of midge-damaged samples from various locations and Mahalanobis classifier gave the highest classification accuracy in classifying the midge-damaged samples (91-100%).

The classification results of samples with crease-up orientation are given in Table 5.34. The LDA, QDA, and Mahalanobis classifiers correctly classified 93.3%, 100%, and 65% healthy kernels, respectively. The classification accuracy of LDA and QDA for classifying midge- damaged kernels varied from low (55%) to very high (100%) and Mahalanobis classifier correctly classified 95.0-100% midge-damaged kernels. Smail et al. (2006) measured the Mahalanobis distance between spectra of sprouted and the mean spectra of healthy kernels by manually selecting 20 pixels in the germ area. The measured Mahalanobis distance also showed promise for use in discriminant analysis. In the discussed classification models, Mahalanobis discriminant classifier gave consistent and highest classification for damaged samples. Both LDA and QDA gave better accuracy in the classification of healthy samples. The overall classification accuracy of

**Table 2.33 Classification of midge-damaged wheat kernels (crease-down) using 300 healthy and 300 midge-damaged kernels from each of five locations (80% training and 20% test sets) (LWNIR)**

<b>Grain type</b>	<b>Classification accuracy (%) of various discriminant classifiers</b>		
Sample	Linear	Quadratic	Mahalanobis
Healthy	96.7	95.0	85.0
Yorkton, SK	86.7	93.3	96.7
Camrose East, AB	100.0	100.0	100.0
Cutknife, SK	96.7	100.0	100.0
Vegreville, AB	83.3	86.7	91.7
North Battleford, SK	100.0	100.0	100.0

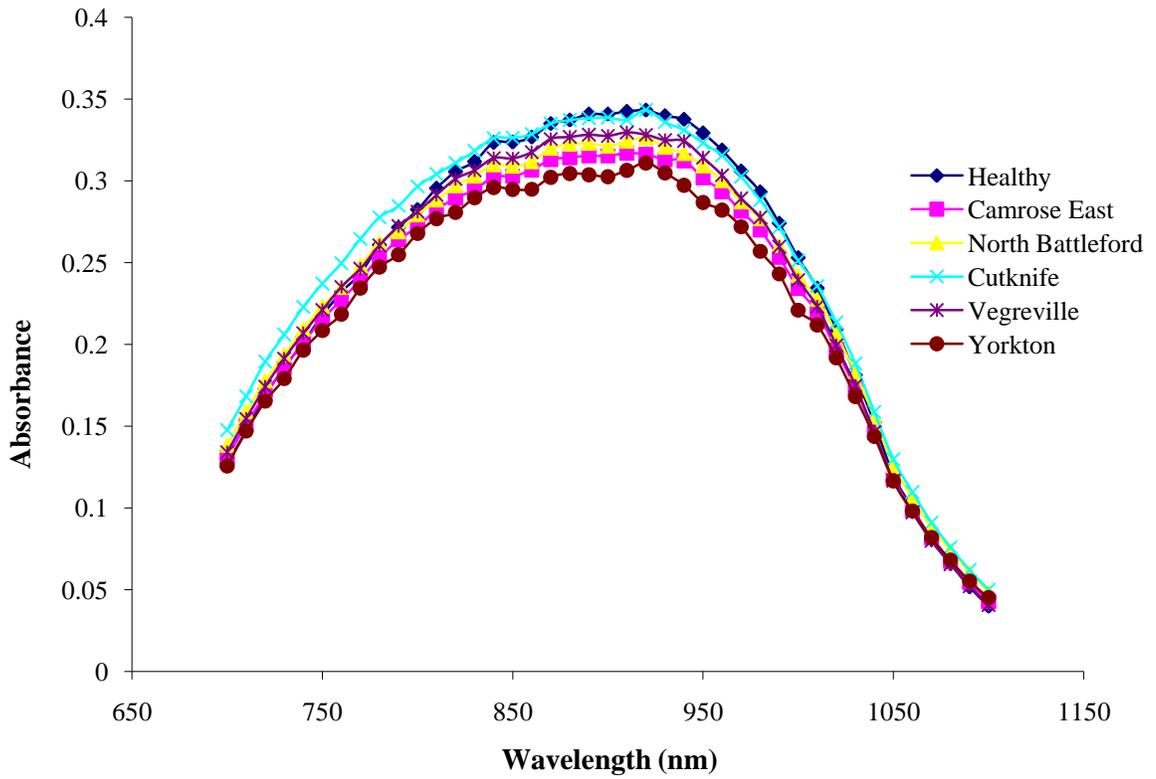
**Table 5.34 Classification of midge-damaged wheat kernels (crease-up) using 300 healthy and 300 midge-damaged kernels from each of five each locations (80% training and 20% test sets) (LWNIR)**

<b>Grain type</b>	<b>Classification accuracy (%) of various discriminant classifiers</b>		
Sample	Linear	Quadratic	Mahalanobis
Healthy	93.3	100.0	65.0
Yorkton, SK	91.7	81.7	98.3
Camrose East, AB	100.0	100.0	100.0
Cutknife, SK	100.0	100.0	100.0
Vegreville, AB	68.3	55.0	95.0
North Battleford, SK	100.0	100.0	100.0

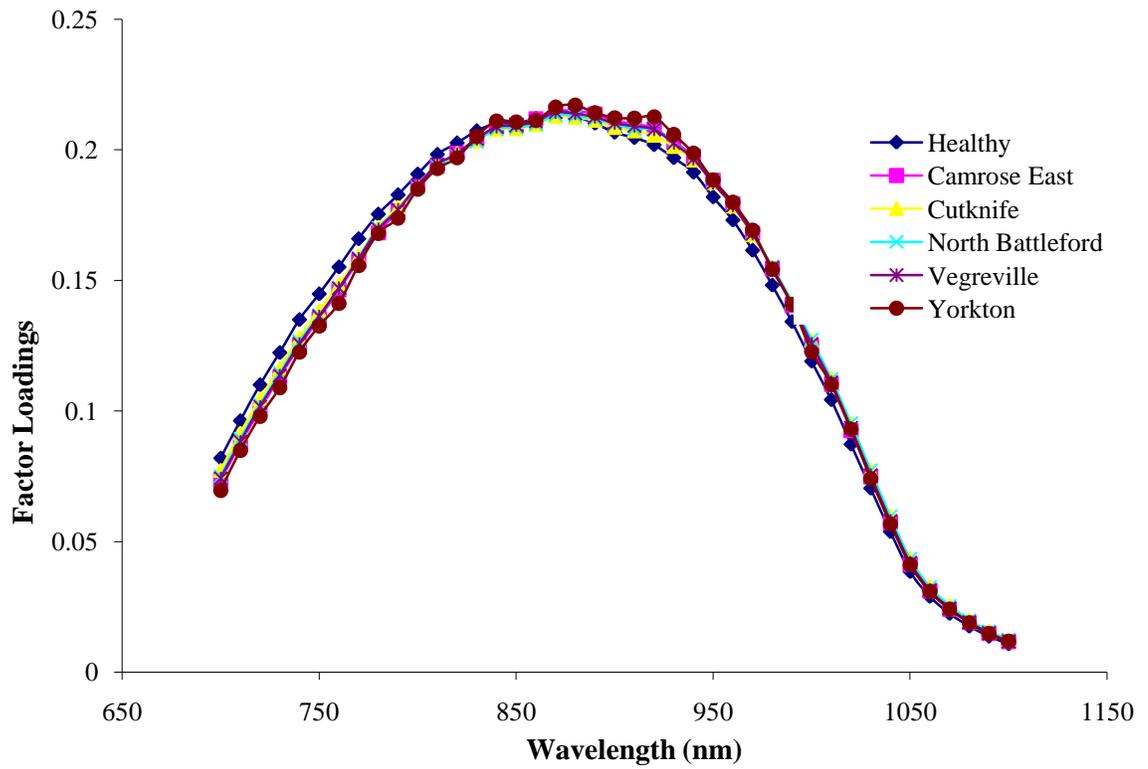
midge-damaged samples by NIR hyperspectral imaging was very high compared to manual inspection (approximately 60%) (Personal communication, Ian Wise, Cereal Research Centre, Winnipeg). The classification accuracy of the crease-down sample was slightly lower (91.67-100%) than crease-up samples (95-100%), however, these results clearly indicate that kernel orientation has a limited effect on the identification of midge-damaged samples compared to visual inspection.

### **5.3.2 Detection of midge-damaged kernels using SWNIR hyperspectral imaging**

Significant wavelengths corresponding to the highest factor loadings of the first PC (above 98.2% variation) from the multivariate image analysis of hyperspectral data were selected. Wavelength of 870 nm was found to be significant which correspond to the CH<sub>3</sub> third overtone region. The absorbance spectra of healthy and midge-damaged kernels in SWNIR region had bell-shaped spectra, whereas, the reflectance or absorbance spectra in LWNIR region had several peaks and valleys. Therefore, only one wavelength was selected as significant wavelength in SWNIR region because addition of another wavelength might cause a co-linearity problem. The NIR absorbance spectra and PC loadings of healthy and midge-damaged wheat kernels are shown in Figs. 5.12 and 5.13, respectively. Two-way statistical classification algorithms were developed using the image features extracted from the significant wavelength image (870 nm). The classification results of healthy and midge-damaged wheat kernels are given in Table 5.35. Midge-damaged wheat kernels were classified with average accuracy of 93.3-98.7% (1.4-6.6 std dev) by Mahalanobis discriminant classifier. Both LDA and QDA correctly classified 87.0-97.6 % (1.9-12.3 std dev) and 85.3-96.0% (2.8-6.8 std dev)



**Fig 5.12 Short-wave near-infrared mean reflectance spectra of healthy and midge-damaged wheat kernels for different locations in western Canada**



**Fig. 5.13 First principal component (PC) mean factor loadings of healthy and midge-damaged wheat kernels (short-wave near-infrared) for different locations in western Canada**

**Table 5.35 Two-way classification of healthy and midge-damaged wheat kernels using SWNIR hyperspectral image features**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	75.0	31.9	86.0	20.4	72.0	34.2
Yorkton, SK	87.0	12.3	85.3	6.8	93.3	6.6
Camrose East, AB	95.0	4.9	96.0	2.8	98.7	1.4
Cutknife, SK	97.6	1.9	93.0	3.2	98.7	1.4
Vegreville, AB	95.0	3.3	87.0	3.6	96.7	1.7
North Battleford, SK	94.3	7.3	93.3	4.9	98.3	2.9

midge-damaged wheat kernels, respectively. However, all three classifiers gave lower classification accuracy in detecting healthy kernels with false positive errors of 14.0-28.0%. The LDA, QDA, and Mahalanobis classifiers correctly classified 75.0 (31.9 std dev), 86.0 (20.4 std dev), and 72.0% (34.2 std dev) healthy kernels, respectively. These classification results showed that SWNIR imaging classification suffered from high false positive errors despite the high midge damage detection rate. The advantage of the SWNIR imaging systems is that the system cost is less compared to the LWNIR imaging systems which mostly uses thermoelectrically cooled InGaAs detectors.

**5.3.3 Detection of midge-damaged kernels using color imaging** Color features are the simplest features in terms of computational complexity and have potentially high discrimination capability. So an attempt was made to use only the top color features to develop a statistically supervised classification algorithm. The top 10 color features ranked using STEPDISC procedure are given in Table 5.36. These color features were

given as input to develop two-way LDA, QDA, and Mahalanobis discriminant classification models. The average classification results of two-way classification using the top 10 color features are given in Table 5.37. The LDA classified 92.7% healthy kernels with 5.6 standard deviation and 87.7-98.7% midge-damaged kernels with standard deviation varying between 1.4-6.3. The QDA classified 86.3% healthy kernels and 91.3-99.3% midge-damaged kernels. The average classification accuracy of Mahalanobis classifier was 78.3% for healthy kernels and 98.0-99.3% for midge-damaged kernels. Both QDA and Mahalanobis classifiers gave high accuracy in classifying midge-damaged kernels and low standard deviation (0.8 to 4.5). However, the classification accuracy of healthy kernels was very low with high standard deviation of 22.2 and 29.9, respectively, for QDA and Mahalanobis classifier.

To further improve the classification performance of the statistical classifiers, two-way classification algorithms were also developed using the top features from all 230 extracted image features. The top 10 features selected from the ranking of the combined 230 features (morphological, color, and textural) by STEPDISC procedure are given in Table 5.38. The first selected feature (gray level run length matrix long run for green band) was a textural feature which demonstrated that textural features carried the highest discriminatory information due to the physical surface damage to the wheat kernels. The top 10 features also included color and morphological features which had distinguishable characteristics as midge-damaged kernels are likely to be darker, shrivelled, and misshapen. The classification algorithms were developed by giving the top 10 image features as input to statistical discriminant classifiers. The classification results of LDA, QDA, and Mahalanobis classifiers are given in Table 5.39. The LDA classified 91.7 %

**Table 5.36 Top 10 features selected from 123 color features of healthy and midge-damaged wheat kernels**

Rank	Feature
1	Green moment 2
2	Green range
3	Red moment 1
4	Hue mean
5	Green moment 2
6	Red moment 2
7	Red mean
8	Blue variance
9	Blue range
10	Green variance

**Table 5.37 Two-way classification of healthy and midge-damaged kernels using top 10 color features from 123 color features extracted from color images of wheat from different locations in western Canada**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	92.7	5.6	86.3	22.2	78.3	29.9
Yorkton, SK	96.0	1.9	98.7	1.4	99.3	1.4
Camrose East, AB	94.0	5.1	94.0	3.7	98.3	2.0
Cutknife, SK	98.7	1.4	99.3	0.9	99.7	0.8
Vegreville, AB	87.7	6.3	91.3	4.5	98.0	1.4
North Battleford, SK	98.3	2.0	98.3	1.7	99.3	1.4

healthy kernels with 17.7% standard deviation and 95.7-99.3% midge-damaged kernels with standard deviation of 0.9-4.0. The QDA classified 93.7% healthy kernels (12.4 std dev) and 94.0-99.7% midge-damaged kernels (0.8-4.8 std dev). Mahalanobis classifier correctly identified 91.0% (18.3 std dev) healthy and 95.3-99.7% midge-damaged kernels (0.8-4.5 std dev). In classifying healthy kernels, using top features from 230 morphological, color, and textural features by QDA and Mahalanobis classifiers, the standard deviation reduced significantly whereas it increased in case of LDA. The QDA classifier gave overall better performance compared to other models. The overall classification accuracy of midge-damaged samples by color imaging was very high compared to manual inspection which was approximately 60% (Personal communication, Ian Wise, Cereal Research Centre, Winnipeg). The classification performance of color imaging was also comparable with the midge-damage classification by NIR hyperspectral imaging (Singh et al., 2009b). Considering the imaging system cost and system simplicity, color imaging systems have higher potential applications in the grain industry.

#### **5.3.4 Detection of midge-damaged kernels using color and SWNIR imaging**

Classification algorithms were developed by combining SWNIR image features with color image features. Firstly, two-way classification models were developed by combining SWNIR features with the top 10 color features (Table 5.40). Midge-damaged kernels were classified with very high accuracy by all three statistical classifiers. The Mahalanobis discriminant classifier correctly identified more than 99.0% midge-damaged kernels (0-1.5 std dev). The LDA and QDA correctly classified 96.3-99.3% (0.9-3.2 std dev) and 97.0-100.0% (0-4.5 std dev) midge-damaged kernels, respectively. Healthy kernels were correctly classified with more than 95.0% (4.0 std dev) accuracy by LDA.

**Table 5.38 Top 10 features selected from 230 morphological, color, and textural features of midge-damaged wheat kernels**

<b>Rank</b>	<b>Feature</b>	<b>Feature type</b>
1	Gray level run length matrix long run for green band	Textural
2	Shape moment 2	Morphological
3	Red moment 1	Color
4	Red histogram range 1	Color
5	Peripheral Fourier descriptor 2	Morphological
6	Hue mean	Color
7	Green moment 1	Color
8	Blue moment 2	Color
9	Gray level run length matrix runpercent for green band	Texture
10	Gray level co-occurrence matrix correlation for blue band	Texture

**Table 5.39 Two-way classification of healthy and midge-damaged wheat kernels using top 10 features from 230 morphological, color, and textural features extracted from color images of wheat from different locations in western Canada**

<b>Sample</b>	<b>Classification by statistical discriminant classifiers (%)</b>					
	<b>Linear</b>		<b>Quadratic</b>		<b>Mahalanobis</b>	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	91.7	17.7	93.7	12.4	91.0	18.3
Yorkton, SK	99.3	1.5	99.7	0.8	99.7	0.8
Camrose East, AB	97.7	2.8	98.0	2.2	98.0	2.2
Cutknife, SK	99.3	0.9	99.0	0.9	99.3	0.9
Vegreville, AB	95.7	4.0	94.0	4.8	95.3	4.5
North Battleford, SK	97.7	2.5	98.3	1.7	98.3	1.7

However, QDA and Mahalanobis classifiers gave very high false positive errors of 14.0-21.1%, respectively, in classifying healthy wheat kernels. To further improve the classification and reduce the false positive error, two-way classification models were also developed by combining SWNIR image features with the top 10 features selected from 230 morphological, color, and textural features derived from the color images (Table 5.41). All three statistical classifiers gave very high classification accuracy of more than 98.0% in detecting midge-damaged wheat kernels. The false positive error of QDA and Mahalanobis discriminant classifier also improved (6-10%). The LDA classifier correctly identified more than 95.0% (5.2 std dev) healthy kernels. The overall performance of LDA was very high with classification rate of 95.3-99.3% (0.9-5.2 std dev) in classifying healthy and midge-damaged wheat kernels. The results showed significant improvement in classification by combining SWNIR image features and digital color image features. These results are also comparable with the classification of LWNIR image features (Singh et al., 2009b).

**Table 5.40 Two-way classification of healthy and midge-damaged wheat kernels using combined top 10 color features from 123 color features and SWNIR hyperspectral image features from different locations in western Canada**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	95.7	4.0	86.0	14.8	74.3	21.1
Yorkton, SK	97.0	3.2	98.3	2.9	99.0	1.5
Camrose East, AB	96.3	0.9	99.3	0.9	99.7	0.74
Cutknife, SK	99.3	0.9	100.0	0.0	100.0	0.0
Vegreville, AB	97.0	3.2	97.0	4.5	99.7	0.74
North Battleford, SK	98.7	1.8	99.7	0.74	99.7	0.8

**Table 5.41 Two-way classification of healthy and midge-damaged wheat kernels using combined top 10 features from 230 morphological, color and textural features and SWNIR hyperspectral image features from different locations in western Canada**

Sample	Classification by statistical discriminant classifiers (%)					
	Linear		Quadratic		Mahalanobis	
	Mean	std dev	Mean	std dev	Mean	std dev
Healthy	95.3	5.2	94.0	6.9	89.7	10.3
Yorkton, SK	99.3	0.9	99.3	0.9	99.3	0.9
Camrose East, AB	98.7	1.8	99.7	0.8	99.7	0.8
Cutknife, SK	99.3	0.9	99.7	0.8	99.7	0.8
Vegreville, AB	98.7	1.4	96.0	6.3	99.3	1.5
North Battleford, SK	98.7	1.4	98.3	1.2	98.7	1.4

## 6. CONCLUSIONS AND RECOMMENDATIONS

The research work conducted in this thesis explored the potential of three imaging systems working in the long-wave NIR (LWNIR) (1000-1600 nm), short-wave NIR (SWNIR) (700-1100 nm), and visible (color) region of the electromagnetic spectrum to classify insect and fungal damage, and sprouted wheat kernels.

In the LWNIR region, wavelengths of 1101.7, 1132.1, 1284.2, 1305.1, 1315.8, and 1347.4 nm were identified as significant and were used in features extraction and classification. The LWNIR hyperspectral images gave high classification accuracy in classifying healthy and damaged kernels and classified 85.0-100.0% healthy and insect-damaged kernels, 95.0-100.0% healthy and fungal-infected kernels, and 85.0-100.0% healthy and sprouted kernels in two-way classification. In most of the LWNIR classification models, the false positive error was below 15.0%. In the SWNIR region, wavelength of 870.0 nm was selected as significant and used in feature extraction and classification. The SWNIR imaging classified 92.7-100.0% insect-damaged kernels, 96.0-100.0% fungal-infected kernels, and 93.3-98.7% sprouted kernels in two-way classification. However, the SWNIR image classification resulted in high false positive error of up to 28.0% and misclassified healthy kernels as damaged.

Classification using the top features from 123 color features classified 88.0-99.0% insect-damaged kernels, 98.0-100.0% fungal-infected kernels, and 98.3-99.7% sprouted kernels in two-way classification with high false positive error (24.0%). The top features selected from 230 morphological, color, and textural features, classified 93.7-99.3% insect-damaged kernels, 98.0-100.0% fungal-infected kernels, and 94.0-99.7% sprouted

kernels in two-way classification but also resulted in a high false positive of up to 26.0% in some cases.

The classification models by combining the top features from 123 color features, and SWNIR features, classified 97.0-100% insect-damaged kernels, 94.3-99.7.0% fungal-infected kernels, and 96.3-99.3% sprouted kernels in two-way classification and reduced false positive error (maximum 17.0%). The classification models developed by combining the top features from 230 morphological, color, and textural features and SWNIR features gave the highest classification accuracy and classified 91.0-100.0% insect-damaged kernels, 99.0-100.0% fungal-infected kernels, and 95.3-99.3% sprouted kernels in two-way classification with less than 4.0% false positive error. The BPNN model did not show any improvement in classification performance over statistical classifiers in preliminary investigation. Statistical classification models using combined SWNIR and digital color image features demonstrated their high potential for real-time applications in grain quality monitoring at grain-handling facilities.

In this thesis, non-touching single kernels were used for feature extraction and classification model development and testing. The performance of the classifiers should also be tested in bulk grain samples. In the grain handling facilities, grain is mostly transferred from one location to an other using a conveyer belt. Therefore, the classification algorithm should be developed to separate touching kernels in bulk grain images in real-time. In the present work, two different systems were used for SWNIR (700-1100 nm) and color image acquisition as in SWNIR system switching from SWNIR filter to VIS filter causes physical movement of filters, requires focus and alignment adjustment and re-calibration. However, in future work, a CCD camera working in the

400-1100 nm wavelength range can be used to scan samples at red, green, blue, and 870 nm wavebands (multispectral) using an interference filter. Red, green, and blue band images can be combined to form RGB composite color images and features can be extracted. The cost of this multispectral imaging system will be significantly lower compared to the research based hyperspectral imaging systems used in our laboratory. Hyperspectral imaging systems are ideal for finding the specific wavelengths for assessing different types of quality determining parameters. These systems use expensive liquid crystal tunable filters (LCTF) that cost over CAD 40,000 for scanning of the samples at hundreds of narrow wavelengths. However, once the specific wavelengths are identified, multispectral imaging systems using less expensive (approx. CAD 1000) bandpass filters (e.g., interference filter) can be developed. Therefore, the real-time multispectral imaging systems will be economical compared to laboratory based hyperspectral imaging research instruments.

## 7. REFERENCES

- AACC International. 2000. Approved Methods of the American Association of Cereal Chemists, 10th Ed. Methods 2208, and 56-81B. St. Paul, MN: AACC.
- Ahmad, I.S., J.F. Reid, and M.R. Paulsen. 1999. Color classifier for symptomatic soybean seeds using image processing. *Plant Disease* 83: 320-327.
- Archibald, D.D., C.N. Thai, and F.E. Dowell. 1998. Development of short-wavelength near-infrared spectral imaging for grain color classification. *Proceedings of SPIE* 3543: 189-198.
- ASAE. 2003. Moisture measurement-unground grain and seeds. In: *ASAE Standards*. St. Joseph, MI: ASAE.
- 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: 80-90.
- Barton, F.E. and D. Burdick. 1979. Preliminary study on the analysis of forages with a filter type near-infrared reflectance spectrometer. *Journal of Agricultural and Food Chemistry* 27: 1248-1252.
- Berardo, N., O.V. Brenna, A. Amatoa, P. Valotia, V. Pisacanea, and M. Motto. 2004. Carotenoids concentration among maize genotypes measured by near infrared reflectance spectroscopy (NIRS). *Innovative Food Science and Emerging Technologies* 5: 393- 398.
- Berman, A, P.M. Connor, L.B. Whitbourn, D.A. Coward, B.G. Osborne, and M.D. Southan. 2007. Classification of sound and stained wheat grains using visible and near infrared hyperspectral image analysis. *Journal of Near Infrared Spectroscopy* 15: 351-358.

- Borjesson, T., U. Stollman, and J. Schnurer. 1993. Off-odorous compounds produced by moulds on oatmeal agar: Identification and relations to other growth characteristics. *Journal of Agricultural Food Chemistry* 41: 2104-2111.
- Brader, B., R.C. Lee, R. Plarre, W. Burkholder, G.B. Kitto, C. Kao, L. Polston, E. Dorneanu, I. Szabof, B. Mead, B. Rouse, D. Sullins, and R. Denning. 2002. A comparison of screening methods for insect contamination in wheat. *Journal of Stored Products Research* 38: 75-86.
- Brodsky, M.H., P. Entis, M.P. Entis, A. Sharpe, and G.A. Jarvis. 1982. Determination of aerobic plate and yeast and mould counts in foods using an automated hydrophobic grid membrane filter technique. *Journal of Food Protection* 45: 301-304.
- Brosnan, T. and D.W. Sun. 2002. Inspection and grading of agricultural and food products by computer vision systems: a review. *Computers and Electronics in Agriculture* 36: 193-213.
- Call, J. and R.A. Lodder. 2002. Application of a liquid crystal tunable filter to near-infrared spectral searches. In *Proceedings of SETI Conference*. Ewing, NJ. April 26-28.
- Canada Grains Act. 1975. Canada grain regulations. In *Canada Gazette*, Part II, Vol. 109, No. 14.
- Casady, W.W., M.R. Paulsen, J.F. Reid, and J.B. Sinclair. 1992. A trainable algorithm for inspection of soybean seed quality. *Transactions of the ASAE* 35(6): 2027-2033.
- CGC. 2006. *Grain Grading Guide*. Winnipeg, MB: Canadian Grain Commission.

- Chambers, J., N.J. McKeivitt, and M.R. Stubbs. 1984. Nuclear magnetic resonance spectroscopy for studying the development and detection of the grain weevil, *Sitophilus granarius* (L.) (Coleoptera: Curculionidae), within wheat kernels. *Bulletin of Entomological Research* 74: 707-724.
- Cheng, X., Y. Tao, Y.R. Chen, and Y. Luo. 2003. NIR/MIR dual-sensor machine vision system for online apple stem-end/calyx recognition. *Transactions of the ASAE* 46(2): 551-558.
- Choudhary, R., J. Paliwal, and D.S. Jayas. 2008. Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images. *Biosystems Engineering* 99(3): 330-337.
- Choudhary, R., S. Mahesh, J. Paliwal, and D.S. Jayas. 2009. Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. *Biosystems Engineering* 102(2): 115-127.
- CIGI. 1993. *Grains & Oilseeds: Handling, Marketing, Processing*, 2th edition. Winnipeg, MB: Canadian International Grains Institute.
- Cogdill, R.P., C.R. Hurburgh, G.R. Rippke, S.J. Bajic, R.W. Jones, J.F. McClelland, T.C. Jensen, and J. Liu. 2004. Single kernel maize analysis by near-infrared hyperspectral imaging. *Transactions of the ASAE* 47(1): 311-320.
- Daftary, R.D. and Y. Pomerranz. 1965. Changes in lipid composition in wheat during storage deterioration. *Journal of Agricultural and Food Chemistry* 13(5): 442-446.

- Davies, E.R., C. Ridgway, and J. Chambers. 2003. NIR detection of grain weevils inside wheat kernels. In *Proceedings of International Conference on Visual Information Engineering*, 173-176. London, UK. July 7-9.
- Delwiche, S.R. 1998. Protein content of single kernel of wheat by near-infrared spectroscopy. *Journal of Cereal Science* 27: 241-254.
- 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 G.A. Hareland. 2004. Detection of scab-damaged hard red spring wheat kernels by near-infrared reflectance. *Cereal Chemistry* 81(5): 643-649.
- Delwiche, S.R. and K.H. Norris. 1993. Classification of hard red wheat by near-infrared diffuse reflectance spectroscopy. *Cereal Chemistry* 70(1): 29-35.
- Delwiche, S.R., Y.R. Chen, and W.R. Hruschkha. 1995. Differentiation of hard red spring wheat by near-infrared analysis of bulk samples. *Cereal Chemistry* 72(3): 243-247.
- Dowell, F.E. 1997. Effect of NaOH on visible wavelength spectra of single wheat kernels and color classification efficiency. *Cereal Chemistry* 74: 617-620.
- Dowell, F.E. 2000. Differentiating vitreous and nonvitreous of durum wheat kernels by using near-infrared spectroscopy. *Cereal Chemistry* 77(2): 155-158.
- Dowell, F.E., J.E. Throne, and J.E. Baker. 1998. Automated nondestructive detection of internal insect infestation of wheat kernels by using near-infrared spectroscopy. *Journal of Economic Entomology* 91: 899-904.

- Dowell, F.E., J.E. Throne, D. Wang, and J.E. Baker. 1999. Identifying stored-grain insects using near-infrared spectroscopy. *Journal of Economic Entomology* 92: 165-169.
- ElMasry, G., N. Wang, A. ElSayed, and M. Ngadi. 2007. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *Journal of Food Engineering* 81: 98-107.
- FAO, 2006. Food and Agriculture Organization of United Organization, Rome.
- FAOSTAT, 2009. Food and Agriculture Organization of United Organization, Rome. <http://faostat.fao.org/> (2009/03/16).
- Fenney, P.L., J.E. Kinney, and J.R. Donelson. 1988. Prediction of damaged starch in straight-grade flour by near-infrared reflectance analysis of whole ground wheat. *Cereal Chemistry* 65: 449-452.
- Gat, N. 2000. Imaging spectroscopy using tunable filters: a review. *Proceedings of SPIE* 4056: 51-64.
- Gaunt, D.M., A.P.J. Trinci, and J.M. Lynch. 1985. The determination of fungal biomass using adenosine triphosphate. *Experimental Mycology* 9: 174-178.
- Geladi, P. and H. Grahn. 1996. *Multivariate Image Analysis*. Chichester, UK: John Wiley and Sons.
- Geladi, P., J. Burger, and T. Lestander. 2004. Hyperspectral imaging: problems and solutions. *Chemometrics and Intelligent Laboratory Systems* 72: 209-217.
- Gendrin, C., Y. Roggo, and C. Collet. 2007. Content uniformity of pharmaceutical solid dosage forms by nearinfrared hyperspectral imaging: A feasibility study. *Talanta* 73(4): 733-741.

- Ghosh, P.K., D.S. Jayas, M.L.H. Gruwel, and N.D.G. White. 2007. A magnetic resonance imaging study of wheat drying kinetics. *Biosystems Engineering* 97(2): 189-199.
- Gordon, S.H., R.B. Schudy, B.C. Wheeler, D.T. Wicklow, and R.V. Greene. 1997. Identification of Fourier transform infrared photoacoustic spectral features for detection of *Aspergillus flavus* infection in corn. *International Journal of Food Microbiology* 35(2): 179-186.
- Gorretta, N., J.M. Roger, M. Aubert, V. Bellon-maurel, F. Campan, and P Roumet. 2006. Determining vitreousness of durum wheat kernels using near infrared hyperspectral imaging. *Journal of Near Infrared Spectroscopy* 14: 231-239.
- Gourama, H. and L.B. Bullerman. 1995. Detection of moulds in foods and feeds: potential rapid and selective methods. *Journal of Food Protection* 58(12): 1389-1394.
- Gowen, A.A., C.P. O'Donnell, P.J. Cullen, G. Downey, and J.M. Frias. 2007. Hyperspectral imaging: an emerging process analytical tool for food quality and safety control. *Trends in Food Science & Technology* 18: 590-598.
- Gowen, A.A., M. Taghizadeh, and C. P. O'Donnell. 2009. Identification of mushrooms subjected to freeze damage using hyperspectral imaging. *Journal of Food Engineering* 93: 7-12.
- Gributs, C.E.W. and D.H. Burns. 2006. Parsimonious calibration models for near-infrared spectroscopy using wavelets and scaling functions. *Chemometrics and Intelligent Laboratory Systems* 83: 44-53.
- Hans, B.P. 2003. Analysis of water in food by near infrared spectroscopy. *Food Chemistry* 82: 107-115.

- Health Canada. 2006. *Policy on Managing Health Risk Associated with the Consumption of Sprouted Seeds and Beans*, 1-8. Ottawa, ON: Health Canada
- Jarvis, B., D.A.L. Sciler, A.J.L. Ould, and A.P. Williams. 1983. Observations on the enumeration of molds in food and feedstuffs. *Journal of Applied Bacteriology* 55:325.
- Kamphuis, H.J., S. Notermans, G.H. Veeneman, J.H. Van-Boom, and F.M. Rombouts. 1989. A rapid and reliable method for the detection of moulds in foods: using latex agglutination assay. *Journal of Food Protection* 52:244-247.
- Karunakaran, C. 2002. Soft X-ray inspection of wheat kernels to detect infestations by stored-grain insects. Unpublished Ph.D. thesis. Winnipeg, MB: Department of Biosystems Engineering, University of Manitoba.
- Karunakaran, C., D.S. Jayas, and N.D.G. White. 2004. Soft X-rays: A potential insect detection method in cereals in grain handling facilities. In *Proceedings of International Quality Grain Conference*. Indianapolis, Indiana. July 19-22.
- Kays, S.E, D.D. Archibald, and M. Sohn. 2005. Prediction of fat in intact cereal food products using near-infrared reflectance spectroscopy. *Journal of the Science of Food and Agriculture* 85: 1596-1602.
- Kim, M.S., Y.R. Chen, and P.M. Mehl. 2001. Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Transactions of the ASAE* 44(3): 721-729.
- Kim, S.S., M.R. Phyu, J.M. Kim, and S.H. Lee. 2003. Authentication of rice using near-infrared reflectance spectroscopy. *Cereal Chemistry* 80(3): 739-745.

- Kim, M.S., A.M. Lefcourt, K. Chao, Y.R. Chen, I. Kim, and D.E. Chan. 2002. Multispectral detection of fecal contamination on apples based on hyperspectral imagery: Part I. Application of visible and near-infrared reflectance imaging. *Transactions of the ASAE* 45(6): 2027-2037.
- Koc, H., V.W. Smail, and D. L. Wetzel. 2008. Reliability of InGaAs focal plane array imaging of wheat germination at early stages. *Journal of Cereal Science* 48(2): 394-400.
- Kruger, J.E. 1994. Enzymes of sprouted grains and possible technological significance. In *Wheat: Production, Properties and Quality*, eds. W. Bushuk and V. Rasper, 143-153. Glasgow, UK: Blackie Academic and Professional.
- Lawrence, K.C., B. Park, W.R. Windham, and C. Mao. 2003a. Calibration of a pushbroom hyperspectral imaging system for agricultural inspection. *Transactions of the ASAE*: 46(2): 513-521.
- Lawrence, K.C., W.R. Windham, B. Park, and R.J. Buhr. 2003b. A hyperspectral imaging system for identification of faecal and ingesta contamination on poultry carcasses. *Journal of Near Infrared Spectroscopy* 11: 269-281.
- Lestander, T. 2003. Multivariate NIR studies of seed-water interaction in scots pine seeds (*Pinus sylvestris L.*). Unpublished Ph.D. thesis. Umea, Sweden: Department of Silviculture, Swedish University of Agricultural Sciences.
- Lewis, N.E., J. Schoppelrei, and E. Lee. 2004. Near-infrared chemical imaging and the PAT initiative. *Spectroscopy* 19(4): 26-35.

- Liu, Y., Y.R. Chen, C.Y. Wang, D.C. Chan, M.S. Kim. 2006. Development of hyperspectral imaging technique for the detection of chilling injury in cucumbers; spectral and image analysis. *Applied Engineering in Agriculture* 22(1): 101-111.
- Lorenz, K. and R. Valvano. 1981. Functional characteristics of sprout-damaged soft white wheat flour. *Journal of Food Science* 46: 1018-1020.
- Lu, R. 2003. Detection of bruise on apples using near-infrared hyperspectral imaging. *Transactions of the ASAE* 46(2): 523-530.
- Lu, R. 2004. Multispectral imaging for predicting firmness and soluble solids content of apple fruit. *Postharvest Biology & Technology* 31: 147-157.
- Lunn, G.D, R.K. Scott, P.S. Kettlewell, and B.J. Major. 1995. Effects of orange wheat blossom midge (*Sitodiplosis mosellana*) infection on pre-maturity sprouting and Hagberg falling number of wheat. *Aspects of Applied Biology* 42: 355-358.
- Madhyastha, S., R.R. Marquardt, and D. Abramson. 1993. Effect of ochratoxin producing fungi on the chemical composition of wheat and barley. *Journal of Food Quality* 16: 287-299.
- Magan, N. 1993. Early detection of fungi in stored grain. *International Biodeterioration & Biodegradation* 32: 145-160.
- Maghirang, E.B. and F.E. Dowell. 2003. Hardness measurement of bulk wheat by single kernel visible and near-infrared reflectance spectroscopy. *Cereal Chemistry* 80(3): 316-322.
- Maghirang, E.B., F.E. Dowell, J.E. Baker, and J.E. Throne. 2003. Automated detection of single wheat kernels containing live or dead insect using near-infrared reflectance spectroscopy. *Transactions of the ASAE* 46, 1277-1282.

- Mahesh, S., A. Manickavasagan, D.S. Jayas, J. Paliwal, and N.D.G. White. 2008. Feasibility of near-infrared hyperspectral imaging to differentiate Canadian wheat classes. *Biosystems Engineering* 101: 50-57.
- Majumdar, S. and D.S. Jayas. 2000a. Classification of cereal grains using machine vision: I. Morphology models. *Transactions of the ASAE* 43(6): 1669-1675.
- Majumdar, S. and D.S. Jayas. 2000b. Classification of cereal grains using machine vision: II. Color models. *Transactions of the ASAE* 43(6): 1677-1680.
- Majumdar, S. and D.S. Jayas. 2000c. Classification of cereal grains using machine vision: III. Texture models. *Transactions of the ASAE* 43(6): 1681-1687.
- Majumdar, S. and D.S. Jayas. 2000d. Classification of cereal grains using machine vision: IV. Morphology, color, and texture models. *Transactions of the ASAE* 43(6): 1689-1694.
- Malone, B.R., C.W. Humphrey, T.R. Romer, and J.L. Richard. 2000. Determination of aflatoxins in grains and raw peanuts by a rapid procedure with fluorometric analysis. *Journal of AOAC International* 83(1): 95-8.
- Manickavasagan, A., D.S. Jayas, and N.D.G. White. 2008. Thermal imaging to detect infestation by *Cryptolestes ferrugineus* inside wheat kernels. *Journal of Stored Products Research* 44: 186-192.
- Manonmani, H.K., S. Anand, A. Chandrashekar, and E.R. Rati. 2005. Detection of aflatoxigenic fungi in selected food commodities by PCR. *Process Biochemistry* 40: 2859-2864.

- Mares, D.J. 1993. Pre-harvest sprouting in wheat: I. Influence of cultivar, rainfall and temperature during grain ripening. *Australian Journal of Agricultural Research* 44:1259-1272.
- 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.
- Muir, W.E. 2001. Production and marketing of cereal and oil seed crops. In: *Grain Preservation Biosystems*. Unpublished manual. Winnipeg, MB: Department of Biosystems Engineering, University of Manitoba.
- Murray, I. and P.C. Williams. 1990. 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.
- Naes, T., T. Isaksson, T. Fearn, and T. Davies. 2002. *A User-Friendly Guide to Multivariate Calibration and Classification*. Chichester, UK: NIR Publications.
- Narvankar, D.S., C.B. Singh, D.S. Jayas, and N.D.G. White. 2009. Assessment of soft X-ray imaging for detection of fungal infection in wheat. *Biosystems Engineering* 103: 49-56.
- Neethirajan, S., D.S. Jayas, and N.D.G. White. 2007. Detection of sprouted wheat kernels using soft X-ray image analysis. *Journal of Food Engineering* 81: 509-513.
- Ng, H.F., W.F. Wilcke, R.V. Morey, and J.P. Lang. 1998. Machine vision evaluation of corn kernel mechanical and mold damage *Transactions of the ASAE* 41(2):415-420.

- Ni, B., M.R. Paulsen, and J.F. Reid. 1998. Size grading of corn kernels with machine vision. *Applied Engineering in Agriculture* 14(5): 567-571.
- Nicolai, B.M., E. Lotze, A. Peirs, N. Scheerlinck, and K.I. Theron. 2006. Non-destructive measurement of bitter pit in apple fruit using hyperspectral imaging. *Postharvest Biology & Technology* 40: 1-6.
- Notermans, S. and H.J. Kamphuis. 1992. Detection of fungi in foods by latex agglutination: a collaborative study. In *Modern Methods of Food Mycology*, eds. R.A. Samson, A.D. Hocking, J.I. Pitt, and A.D. King, 205-212. New York, NY: Elsevier.
- Notermans, S., C.J. Heuvelman, R.R. Beumer, and R. Maas. 1986. Immunological detection of moulds in foods relation between antigen production and growth. *International Journal of Food Microbiology* 3: 253-261.
- Oakley, J.N. 1994. Orange wheat blossom midge: A literature review and survey of 1993 outbreak. In *HGCA Research Review*, No. 28, pp 51. London, UK: Home Grown Cereal Authority.
- Oakley, J.N., P.C. Cumbleton, S.J. Corbett, P. Saunderst, D.I. Green, J.E.B. Young, and R. Rodgers. 1998. Prediction of orange wheat blossom midge activity and risk of damage. *Crop Protection* 17(2): 145-149.
- Osborne, B.G., T. Fearn, and P.H. Hindle. 1993. Theory of near-infrared spectrometry. In *Near Infrared Spectroscopy in Food Analysis*. Singapore: Longman Singapore Publishers.

- Paliwal, J. 2002. Digital image analysis of grain samples for potential use in grain cleaning. Unpublished Ph.D. thesis. Winnipeg, MB: Department of Biosystems Engineering, University of Manitoba.
- Paliwal, J., N.S. Visen, D.S. Jayas, and N.D.G. White. 2003. Cereal grains and dockage identification using machine vision. *Biosystems Engineering* 85(1): 51-57.
- Paliwal, J., 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.
- Paolesse, R., A. Alimelli, E. Martinelli, C.D. Natale, A. D'Amico, M.G. D'Egidio, G. Aureli, A. Ricelli, and C. Fanelli. 2006. Detection of fungal contamination of cereal grain samples by an electronic nose. *Sensors and Actuators B* 119: 425-430.
- Pasikatan, M.C. and F.E. Dowell. 2004. High-speed segregation of high and low-protein single wheat seeds. *Cereal Chemistry* 81(1): 145-150.
- Pasquini, C. 2003. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of Brazilian Chemical Society* 14(2): 198-219.
- Paulsen, M. R. 1990. Using machine vision to inspect oilseeds. *Inform* 1: 50-55.
- Pearson, T.C. and D.L. Brabec. 2007. Detection of wheat kernels with hidden insect infestations with an electrically conductive roller mill. *Applied Engineering in Agriculture* 23: 639-645.
- Pearson, T.C. and D.T. Wicklow. 2006. Detection of corn kernels infected by fungi. *Transactions of the ASAE* 49(4): 1235-1245.

- Pearson, T.C., A.E. Cetin, A.H. Tewfik, and R.P. Haff. 2007. Feasibility of impact-acoustic emissions for detection of damaged wheat kernels. *Digital Signal Processing* 17: 617-633.
- Pearson, T.C., D.T. Wicklow, E.B. Maghirang, F. Xie, and F.E. Dowell. 2001. Detecting aflatoxin in single corn kernels by using transmittance and reflectance spectroscopy. *Transactions of the 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.
- Ram, M.S., F.E. Dowell, L. Seitz, and G. Lookhart. 2002. Development of standard procedures for a simple, rapid test to determine wheat color class. *Cereal Chemistry* 79(2): 230-237.
- Ridgeway, C. and J. Chambers. 1998. Detection of insects inside wheat kernels by NIR imaging. *Journal of Near-Infrared Spectroscopy* 6: 115-119.
- Ridgeway C., J. Chambers, and I.A. Cowe. 1999. Detection of grain weevils inside single wheat kernels by a very near infrared two-wavelength model. *Journal of Near-Infrared Spectroscopy* 7: 213-221.
- Ridgeway, C., E.R. Davies, J. Chambers, D.R. Mason, and M. Bateman. 2002. Rapid machine vision method for the detection of insects and other particulate bio-contaminants of bulk grain in transit. *Biosystems Engineering* 83(1): 21-30.
- Rossi, T.M. and I.M. Warner. 1986. Rank estimation of emission excitation matrices using frequency analysis of eigenvectors. *Analytical Chemistry* 58: 810-815.

- Sanchez-Marinez, R.I., M.O. Cortez-Rocha, F. Ortega-Dorame, M. Morales-Valdes, and M.I. Silveira. 1997. End-use quality of flour from *Rhizopertha dominica* infested wheat. *Cereal Chemistry* 74: 481-483.
- Sashidhar R.B., R.V. Sudershan, Y. Ramkrishna, S. Nahdi, and R.V. Bhat. 1988. Enhanced fluorescence of ergosterol by iodination and determination of ergosterol by fluorodensitometry. *Analyst* 113(5): 809-812.
- Sasic, S. 2000. Eigenvalues and principal component loadings of heavily overlapped vibrational spectra. *Spectrochimica Acta* A57: 323-336.
- Scotter, J.M., V.C. Langford, F.P. Wilson, M.J. McEwan, and S.T. Chambers. 2005. Real-time detection of common microbial volatile organic compounds from medically important fungi by selected ion flow tube-mass spectrometry (SIFT-MS). *Journal of Microbiological Methods* 63: 127-134.
- Seitz, L.M., D.B. Sauer, R. Burroughs, H.E. Mohr, and J.D. Hubbard. 1979. Ergosterol as a measure of fungal growth. *Phytopathology* 69: 1202-1203.
- Seitz, L.M., H.E. Mohr, R. Burroughs, and D.B. Sauer. 1977. Ergosterol as an indicator of fungal invasion in grains. *Cereal Chemistry* 54: 1207-1217.
- Shahin, M.A and S.J. Symons. 2008. Detection of hard vitreous and starchy kernels in amber durum wheat samples using hyperspectral imaging. *NIR News* 19(5): 16-18.
- Shashikumar, K., J.L. Hazelton, G.H. Ryu, and C.E. Walker. 1993. Predicting wheat sprout damage by near-infrared reflectance analysis. *Cereal Foods World* 38(5): 364-366.

- Singh, C.B., D.S. Jayas, J. Paliwal, and N.D.G. White. 2007. Fungal detection in wheat using near-infrared hyperspectral imaging. *Transactions of the ASABE* 50(6): 2171-2176.
- Singh, C.B., D.S. Jayas, J. Paliwal, and N.D.G. White. 2009a. Detection of insect-damaged wheat kernels using near-infrared hyperspectral imaging. *Journal of Stored Products Research* 45(3): 151-158.
- Singh, C.B., J. Paliwal, D.S. Jayas, and N.D.G. White. 2006. Near-infrared spectroscopy: Applications in the grain industry. CSBE Paper No. 06-189, Winnipeg, MB: CSBE.
- Singh, C.B., D.S. Jayas, J. Paliwal, and N.D.G. White. 2009b. Detection of sprouted and midge-damaged wheat kernels using near-infrared hyperspectral imaging. *Cereal Chemistry* 86(3): 256-260.
- Slaughter, D.C., K.H. Norris, and W.R. Hruschka. 1992. Quality analysis and classification of hard red winter wheat. *Cereal Chemistry* 69: 428-432.
- Smail, V.W., A. K. Fritz, and D. L. Wetzel. 2006. Chemical imaging of intact seeds with NIR focal plane array assists plant breeding. *Vibrational Spectroscopy* 42: 215-221.
- Steenhoek, L.W., M.K. Misra, W.D. Batchelor, and J.L. Davidson. 2001. Probabilistic neural networks for segmentation of features in corn kernel images. *Applied Engineering in Agriculture* 17(2): 225-234.
- Sukurai, Y. and H. Shiota. 1977. On the convenient method for glucosamine estimation in Koji. *Agricultural and Biological Chemistry* 41: 619-624.

- Tatzer, P., M. Wolf, and T. Panner. 2005. Industrial application for inline material sorting using hyperspectral imaging in the NIR range. *Real-Time Imaging* 2: 99-107.
- Toews, M.D., J. Perez-Mendoza, J.E. Throne, F.E. Dowell, E. Maghirang, F.H. Arthur, and J.F. Cambell. 2007. Rapid assessment of insect fragments in flour milled from wheat infested with known densities of immature and adult *Sitophilus oryzae* (Coleoptera: Curculionidae). *Journal of Economic Entomology* 100: 1704-1723.
- Toews, M.D., T.C. Pearson, and J.F. Campbell. 2006. Imaging and automated detection of *Sitophilus oryzae* L. (Coleoptera: Curculionidae) pupae in hard red winter wheat. *Journal of Economic Entomology* 99: 583-592.
- Tran, C.D. 2003. Infrared multispectral imaging: principle and instrumentation. *Applied Spectroscopy Reviews* 38(2): 133-153.
- Tseng, T.C., J.C. Tu, and S.S. Tzean. 1995. Mycoflora and mycotoxins in dry bean (*Phaseolus vulgaris*) produced in Taiwan and in Ontario, Canada. *Botanical Bulletin of Academia Sinica* 36: 229-234.
- Vadivambal, R., D.S. Jayas, and N.D.G. White. 2007. Wheat disinfestation using microwave energy. *Journal of Stored Products Research* 43(4): 508-514.
- Van der Maaten, L.J.P. 2007. *An Introduction to Dimensionality Reduction Using Matlab*. Report No., MICC 07-07. Maastricht, the Netherlands: Maastricht University.
- Visen, N.S. 2002. Machine vision based grain handling system. Unpublished Ph.D. thesis. Winnipeg, MB: Department of Biosystems Engineering, University of Manitoba.

- Walcott, R.R. and D.C. McGee. 1998. Detection of asymptomatic fungal infections of soybean seeds by ultrasound analysis. *Plant Disease* 82: 584-589.
- Wan, Y.N., C.M. Lin, and J.F. Chiou. 2002. Rice quality classification using an automatic grain quality inspection system. *Transactions of the ASAE* 45(2): 379-387.
- 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., F.E. Dowell, and R.E. Lacey. 1999. Single kernel color classification by using near-infrared reflectance spectra. *Cereal Chemistry* 76(1): 30-33.
- Wang, D., F.E. Dowell, M.S. Ram, and W.T. Schapaugh. 2003. Classification of fungal-damaged soybean seeds using near-infrared spectroscopy. *International Journal of Food Properties* 7: 75-82.
- Wang, W. and J. Paliwal. 2006. Spectral data compression and analyses techniques to discriminate wheat classes. *Transactions of the ASABE* 49(5): 1607-1612.
- Wang, W. and J. Paliwal. 2007. Near-infrared spectroscopy and imaging in food quality and safety. *Sensing and Instrumentation for Food Quality and Safety* 1(4): 193-207.
- 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* 34: 125-133.

- White, G.G. and T.A. Lambkin. 1988. Damage to wheat grain by larvae of *Tribolium castaneum* (Herbst) (Coleoptera: Tenebrionidae). *Journal of Stored Products Research* 24: 61-65.
- White, N.D.G. 1995. Insects, mites, and insecticides in stored grain ecosystems. In *Stored-Grain Ecosystems*, eds. Jayas, D.S., N.D.G. White, and W.E. Muir, 123-168. New York, NY: Marcel Dekker Inc.
- Williams, A.P. 1989. Methodology developments in food mycology. *Journal of Applied Bacteriology* 67(S): 612-617.
- Xie, F., F.E. Dowell, and X.S. Sun. 2004. Using visible and near-infrared spectroscopy and calorimetry to study starch, protein and temperature effects on bread stalling. *Cereal Chemistry* 81(2): 249-254.
- Yao, H., Z. Hruska, K. DiCrispino, K. Brabham, D. Lewis, J. Beach, R.L. Brown, and T.E. Cleveland. 2005. Differentiation of fungi using hyperspectral imagery for food inspection. ASAE Paper No. 053127. St. Joseph, MI: ASAE.
- Zayas, I.Y. and P.W. Flinn. 1998. Detection of insects in bulk wheat samples with machine vision. *Transactions of the ASAE* 41(3): 883-888.
- Zeringue, H.J. Jr., D. Bhatnagar, and T.E. Cleveland. 1993. C<sub>15</sub>H<sub>24</sub> volatile compounds unique to aflatoxigenic strains of *Aspergillus flavus*. *Applied Environmental Microbiology* 59: 2264-2270.
- Zeringue, H.J.Jr., B.Y. Shih, K. Maskos, and D. Grimm. 1999. Identification of the bright-greenish-yellow-fluorescence (BGY-F) compound on cotton lint associated with aflatoxin contamination in cottonseed. *Phytochemistry* 52(8): 1391-1397.

Zhang, H. and J. Wang. 2007. Detection of age and insect damage incurred by wheat, with an electronic nose. *Journal of Stored Products Research* 43: 489-495.

Zhang, M., L.I. Ludas, M.T. Morgan, G.W. Kurtz, and C.J. Percetti. 1999. Applications of color machine vision in the agricultural and food industries. *Proceedings of SPIE* 3543: 208-219.

## **8. APPENDIX**

**Table A.1 Applications of NIR spectroscopy in grain quality analysis**

Sl. No.	Product	Objective	Wavelength range (nm)	Pre-treatment	Data analysis	Reference
1	Wheat	Cultivar Classification	551-750, 1120-2476	2 <sup>nd</sup> difference, log transform	PLS, MLR	Delwiche and Masie (1996)
2	Wheat	Color classification	500-1900	1 <sup>st</sup> and 2 <sup>nd</sup> derivative	PLS,	Wang et al. (1999) Ram et al. (2002)
3	Cereals	Classification	400-1850	Savitzky-Golay smoothing	BPNN	Mohan et al. (2005)
4	Wheat	Growing season, hardness	1100-2498 400-1700	2 <sup>nd</sup> derivative	MLR, PCA, PLS, ANN,	Delwiche and Norris (1993), Delwiche et al. (1995), Maghirang and Dowell (2003)
5	Wheat	Vitreousness	400-1700	Mean centering	PLS, DA	Dowell (2000) Wang et al. (2002)
6	Wheat	Insect	400-2500 another range	MSC, SNV	PLS, ANN, PCA, PLS	Maghirang et al. (2003) Paliwal et al. (2004),
7	Soybean	Fungal damage	490-1690	-	PLS, ANN	Wang et al. (2003)
8	Corn	Mycotoxins	500-1700	Savitzky-Golay Smoothing, spectral ratio	PLS, DA	Pearson et al. (2001)
9	Wheat	Moisture	1100-2500	Wavelet transform	GA	Gribtus and Burns (2006)
10	Corn	Carotenoid	400-2498	-	PLS	Berardo et al. (2004)
11	Wheat	Protein composition	400-2498	-	PLS	Wesley et al. (2001)
12	Cereal Foods	Fat Content	1100-1700	MSC , derivative	PLS	Kays et al. (2005)

**Table A.2 Applications of imaging processing in grain quality analysis**

<b>Sl. No.</b>	<b>Grain type</b>	<b>Objective</b>	<b>Pre-processing</b>	<b>Input image parameter</b>	<b>Classification method</b>	<b>Reference</b>
1	Corn	Surface analysis	Histogram thresholding	Color features	PNN	Steenhoek et al. (2001)
2	Corn	Size grading	Fixed thresholding	Morphological features	LDA	Ni et al (1998)
3	Corn	Mechanical and mold damage	G/R ratio thresholding	Color features	BPNN	Ng et al. (1998)
4	Rice	Appearance characteristics	Histogram thresholding	Color, morphological features	Logical “and”	Wan et al. (2002)
5	Wheat	Insect detection	Subimage selection,	Color features	DA	Zayas and Flinn (1998)
6	Cereal grains	Classification	Automatic thresholding	Morphological, color, textural features	DA	Majumdar and Jayas (2000a-2000d)
7.	Cereal grains	Dockage identification	-	Morphological, color, textural features	BPNN	Paliwal et al. (2003)
8.	Cereal grains	Classification	-	Wavelet, morphological, color, textural features	LDA, QDA	Choudhary et al. (2008)

**Table A.3 Reported applications of hyperspectral imaging**

Sl No.	Product	Analysis	Wavelength range (nm)	Preprocessing	Classification method	Reference
1.	Tablets	API content, distribution	1200-2400, 1100-2440	Mean filtering, SNV, smoothing (SG)	PLS, CLS	Lewis et al. (2004), Gendrin et al. (2007)
2.	Paper	Quality	900-1700	Dark and light pixel removal	PCA, LDA, KNN	Tatzer et al. (2005)
3.	Strawberry	Quality attribute	400-1000	Binary mask, MSC, SG	PLS, MLR	ElMasry et al. (2007)
4.	Apple	Surface Contamination	450-851	Masking	PCA	Kim et al. (2002)
5.	Apple	Bruise detection	900-1700	Filtering, stretching	PCA, MNF	Lu (2003)
6.	Cucumber	Chilling injury	450-950	MSC, MC, SG masking	PCA, band-ratio	Liu et al. (2006)
7.	Poultry	Contamination	400-2498, 430-900	MSC, masking	PCA, image ratios	Lawrence et al. (2003b)
8.	Wheat	Vitreousness	650-1100	Erosion, SNV, SG smoothing,	PLS-FDA	Gorretta et al. (2004)
9.	Maize	Moisture & protein	750-1090	SNV, DT, MSC	GA, PCR, PLS	Cogdill et al. (2004)
10	Wheat	Classification	1000-1700	Image cropping	DA, BPNN	Mahesh et al. (2008)
11	Wheat	Vitreousness	900-2500	SG smoothing	-	Shahin and Symons (2008)
12	Wheat	Classification	1000-1600	Wavelet	DA, BPNN	Choudhary et al. (2009)
13	Mushroom	Freeze damage	400-1000	SNV	PCA, LDA	Gowen et al. (2009)

**Table A4. Four-way classification of healthy and fungal-infected kernels by LDA using SWNIR image features**

	Healthy	<i>Penicillium</i>	<i>A. glaucus</i>	<i>A. niger</i>
Healthy	<b>85</b>	6.7	6.7	1.6
<i>Penicillium</i>	0	<b>78.3</b>	16.7	5
<i>A. glaucus</i>	1.6	1.7	<b>91.7</b>	5
<i>A. niger</i>	0	0	1.7	<b>98.3</b>

**Table A5. Four-way classification of healthy and fungal-infected kernels by QDA using SWNIR image features**

	Healthy	<i>Penicillium</i>	<i>A. glaucus</i>	<i>A. niger</i>
Healthy	<b>86.7</b>	3.3	6.7	3.3
<i>Penicillium</i>	1.6	<b>76.7</b>	16.7	5
<i>A. glaucus</i>	0	1.7	<b>98.3</b>	0
<i>A. niger</i>	0	0	11.7	<b>88.3</b>

**Table A6. Four-way classification of healthy and fungal-infected kernels by Mahalanobis classifier using SWNIR image features**

	Healthy	<i>Penicillium</i>	<i>A. glaucus</i>	<i>A. niger</i>
Healthy	<b>100</b>	0	0	0
<i>Penicillium</i>	8.3	<b>71.7</b>	13.3	6.7
<i>A. glaucus</i>	5	1.7	<b>90.0</b>	3.3
<i>A. niger</i>	1.6	0	1.7	<b>96.7</b>