UC Davis

UC Davis Previously Published Works

Title

Importance of spatial and spectral data reduction in the detection of internal defects in food products.

Permalink

<https://escholarship.org/uc/item/8hh6f6w2>

Journal

Applied spectroscopy, 69(4)

ISSN

0003-7028

Authors

Zhang, Xuechen
Nansen, Christian
Aryamanesh, Nader
et al.

Publication Date

2015-04-01

DOI

10.1366/14-07672

Peer reviewed

Importance of Spatial and Spectral Data Reduction in the Detection of Internal Defects in Food Products

Xuechen Zhang,^{a,b,c,f} Christian Nansen,^{a,c,d,*} Nader Aryamanesh,^{b,c} Guijun Yan,^{b,c} Farid Boussaid^e

^a University of Western Australia, School of Animal Biology, Faculty of Science, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia

^b University of Western Australia, School of Plant Biology, Faculty of Science, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia

^c University of Western Australia, UWA Institute of Agriculture, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia

^d University of California—Davis Department of Entomology and Nematology, Briggs Hall, Room 367, Davis, CA 95616, USA

^e University of Western Australia, School of Electrical, Electronic and Computer Engineering, 35 Stirling Highway, Crawley, Perth, WA 6009, Australia

^f University of Tasmania, P.O. Box 46, Kings Meadows TAS 7249, Australia

Despite the importance of data reduction as part of the processing of reflection-based classifications, this study represents one of the first in which the effects of both spatial and spectral data reductions on classification accuracies are quantified. Furthermore, the effects of approaches to data reduction were quantified for two separate classification methods, linear discriminant analysis (LDA) and support vector machine (SVM). As the model dataset, reflection data were acquired using a hyperspectral camera in 230 spectral channels from 401 to 879 nm (spectral resolution of 2.1 nm) from field pea (*Pisum sativum*) samples with and without internal pea weevil (*Bruchus pisorum*) infestation. We deployed five levels of spatial data reduction (binning) and eight levels of spectral data reduction (40 datasets). Forward stepwise LDA was used to select and include only spectral channels contributing the most to the separation of pixels from non-infested and infested field peas. Classification accuracies obtained with LDA and SVM were based on the classification of independent validation datasets. Overall, SVMs had significantly higher classification accuracies than LDAs ($P < 0.01$). There was a negative association between pixel resolution and classification accuracy, while spectral binning equivalent to up to 98% data reduction had negligible effect on classification accuracies. This study supports the potential use of reflection-based technologies in the quality control of food products with internal defects, and it highlights that spatial and spectral data reductions can (1) improve classification accuracies, (2) vastly decrease computer constraints, and (3) reduce analytical concerns associated with classifications of large and high-dimensional datasets.

Index Headings: Machine vision; Spatial resolution; Spectral resolution; Pea weevil; Field pea; Reflection data; Hyperspectral imaging.

INTRODUCTION

Hyperspectral imaging technology has been widely used for the detection of insect-induced damage to food products, including field peas (*Pisum sativum*),¹ wheat kernels (*Triticum aestivum*),^{2,3} soybeans (*Glycine max*),⁴ and jujubes (*Ziziphus jujuba*).^{5,6} The common denominator in these reflection-based applications is to accu-

ately and consistently detect changes in reflection profiles and associate such changes with damage or loss in food quality. Because insect-induced damage often causes only subtle changes in the reflection profiles acquired from food products^{1–6} and crops,^{7,8} the use of sensitive, reliable, and robust classification methods is of paramount importance. A support vector machine (SVM) seeks a decision boundary, providing a trade-off between fitting the training data and hypothesis space complexity.⁹ Widely used to classify reflection data from agricultural applications of high-resolution remote sensing, SVM is used for such tasks as the detection of bruises on red bayberry (*Myrica cerifera*),¹⁰ detection of plant diseases,¹¹ classification of near isogenic maize lines (*Zea mays*),¹² and discrimination between black walnut (*Juglans nigra*) shell and pulp.¹³

To optimize reflection-based classifications of food products, hyperspectral imaging data are typically acquired at high spatial (several pixels per square millimeter) and spectral resolutions (spectral channels covering wavelength ranges less than 10 nm).^{14,15} However, when SVM classification is used, such datasets often pose a significant challenge because they become quite large and therefore imply considerable requirements regarding data-processing capacity.^{15–17} Thus, to effectively use SVM, we must determine ways to reduce the size of the input datasets. Data reduction can be achieved at the spatial level through spatial binning, in which reflection profiles across neighboring pixels are averaged.^{15,18} For spectral data reduction, there are three approaches: (1) reduce the number of spectral channels used as explanatory variables¹⁹ through forward stepwise selection of spectral channels with the highest level of contribution, (2) combine effects of a reduced number of individual channels into a small number of factors using partial least square or similar methods,²⁰ and (3) increase the width or range of spectral channels (and decrease spectral resolution) using spectral binning of adjacent spectral channels.^{14,15} In addition to faster data processing, the reduction of spatial and spectral data may reduce the levels of stochastic noise and levels of autocorrelation among adjacent spectral channels and, therefore, increase classification accuracy.²¹ In addition, data reduction

Received 31 July 2014; accepted 30 October 2014.

* Author to whom correspondence should be sent. E-mail: chrnansen@ucdavis.edu.

DOI: 10.1366/14-07672

may reduce the risk of model overfitting due to the Hughes phenomenon^{10,22} and the risk of overfitting or violation of the principle of parsimony.²³ Nevertheless, excessive spatial and spectral data reduction may cause important radiometric information to be lost or masked. With the commercial interest in both fast data processing and also robust and accurate data classification (i.e., based on SVM), it is important to assess the consequences of data reduction on the accuracy of classification models.

Linear discriminant analysis (LDA) is a simple and efficient algorithm for the classification of two or more classes of target objects.²⁴ Compared to SVM, LDA requires less time for analysis because no parameter tuning is required.¹⁶ In addition, LDA is a powerful tool for reducing the dimensionality of hyperspectral imaging input data to a smaller number of dimensions,²⁵ and it can be used, through forward stepwise selection, to select and include the spectral channels with highest contributions to the classification. Thus, data reduction through the selection of only the “best” spectral channels may, in addition to reducing the amount of input data, increase the classification accuracy and make the classification model less sensitive to noise and stochasticity. Linear discriminant analysis has been widely used in hyperspectral imaging analysis on data classification and dimensionality reduction, such as the early detection of apple bruises,²⁶ detection insect-damaged wheat,^{2,3} classification of infected tobacco leaves,²⁵ identification of freeze-damaged mushrooms,²⁷ and detection of bonemeal fragments in animal feed samples.²⁸

Field pea (*P. sativum*) is the second-most important cultivated grain legume in the world.²⁹ Field peas are produced for both human consumption and animal feed, and they contain a high protein content (18–30% among cultivars) and a relatively high content of lysine, suggesting they are a good dietary supplement to cereals for human beings.³⁰ The pea weevil (*Bruchus pisorum*) is one of the most damaging pests of cultivated field peas in Australia,³¹ Europe,³² and the Americas.³³ Similar to damage caused by other bruchids and true weevils (Curculionidae), females lay eggs on the surface of immature pea pods, and once a neonate larva reaches the inside of a pod, it feeds and develops inside an individual field pea.³¹ Heavily weevil-infested field peas show significant weight loss and cannot be used for human consumption. After completing larval development, the mature weevil exits a pea through a 2–3 mm diameter hole. The separation of field peas into those with and without exit holes through visual inspection is time consuming and labor intensive, and it does not eliminate weevil-infested field peas without exit holes. A novel 30% cesium chloride (CsCl)-based density separation method can be used effectively (infested field peas float and non-infested field peas sink).³⁴ However, field peas immersed in 30% CsCl cannot be used for consumption (only for plant-breeding programs).

Despite the importance of data reduction as part of the processing of reflection-based classifications, this study represents one of the first that quantifies the effects of both spatial and spectral data reductions on classification accuracy for both SVM and LDA classification

models. We acquired reflection data from a food product with and without internal defects using a hyperspectral camera in 230 spectral channels from 401 to 879 nm (spectral resolution of 2.1 nm). We deployed five levels of spatial data reduction (binning): no spatial binning (50,000 pixels/mm²), binning pixels into 2 × 2 grids (12,500 pixels/mm²), binning pixels into 5 × 5 grids (2,000 pixels/mm²), 10 × 10 grids (0.500 pixel/mm²), and 20 × 20 grids (0.125 pixel/mm²). For each of the five spatial data reductions, we deployed eight levels of spectral data reduction by binning reflection values across the following number of spectral channels: 230 (no spectral binning), 115, 77, 57, 46, 23, 10, and 5. For each of the 40 datasets (five spatial resolutions times eight spectral resolutions), stepwise LDA was used to select and include spectral channels contributing the most to the separation of non-infested from infested field peas. Based on the selected spectral channels, we conducted LDA and SVM classifications and calculated classification accuracies based on independent validation data (reflection data acquired from field pea samples not included in the training dataset).

MATERIALS AND METHODS

Field Pea Samples. Five commercial field pea cultivars, which are grown mainly for human consumption (Pennant, Dunwa, Helena, Kasper, and Yarrum) and seven pea weevil-resistant field pea backcross lines (BC₁F₆)^{34,35} were included as the training data in this study; their average reflection profiles are presented in Fig. 1a. This selection of field pea cultivars represented a wide range of typical background colors of field peas (Fig. 2). Infested field pea samples were obtained by using an existing pea weevil colony to infest peas under laboratory conditions. After infestation, individual field peas were selected using the CsCl separation method³⁴ to accurately divide the field peas into infested and non-infested classes. A total of 24 classes (12 pea samples with weevil infestation and 12 samples without weevil infestation) were used as the training data. For each class, we imaged individual field peas (5–10 in each subsample) with three replications, and we used different field peas in each replication. Despite the variation in background colors of the different field pea cultivars, pea weevil infestation caused a consistent darkening (reduced reflection) in most of the field pea samples across the examined wavelength spectrum (Fig. 1b). Thus, a reflection-based classification approach was supported by qualitative visual inspection.

As independent validation samples, we acquired additional reflection data from the five commercial cultivars (Pennant, Dunwa, Helena, Kasper, and Yarrum). These validation data were acquired from samples without pea weevil infestation and from samples of infested field peas without external holes (five varieties times three replications with and without weevil infestations, totaling 30 validation samples). Reflection data from all the validation samples were acquired on a separate day than those used as training data.

Hyperspectral Image Collection. We used a hyperspectral spectral camera (PIKA II; Resonon Inc., Bozeman, MT) mounted 40 cm above the field peas. The

FIG. 1. (a) Average reflection profiles acquired from non-infested field peas. (b) Average reflection profiles acquired from infested field peas divided by average reflection profiles acquired from non-infested field peas.

hyperspectral imaging data acquisition was the same as described in previously published studies.^{28,36,37} Hyperspectral images were acquired with the spatial resolution of 50 pixels/mm². Hyperspectral images were collected under artificial lighting from two 15 W, 12 V light bulbs mounted on either side of the lens. During the experiment, the relative humidity was between 30 and 40%, and the temperature was between 19 and 22 °C in the lab. The dark calibration (reflection of 0) was performed by covering the imaging lens, and the white calibration (reflection of 1) was performed using a piece of white Teflon (K-Mac Plastics, USA) and was conducted

for the entire field of view, therefore accounting for possible spatial variability in light conditions. Consequently, the reflection values in all channels ranged between 0 and 1. Each hyperspectral image cube consisted of 100 frames (64 000 pixels), and the hyperspectral spectral camera acquired reflection data in 240 spectral channels from 392 to 889 nm (spectral resolution of 2.1 nm). The native spectral sampling resolution of the instrument was 7.4 nm, and the full width at half-maximum (FWHM) was 240 spectral channels across 500 nm (400–900 nm) for 2.1 nm/channel. We omitted the five spectral channels at either end of the sampled

FIG. 2. Examples of field peas included in this study.

Fig. 3. Representative presentation of a field pea at the five spatial binnings.

spectrum (392–889 nm) because these were associated with stochastic noise. Consequently, only 230 spectral channels from 401 to 879 nm were included. Colored plastic cards (green, yellow, and red) were imaged at the imaging event, and the average reflection profiles from these cards were used to confirm the high consistency of the hyperspectral image acquisition conditions (less than 2% variance in individual spectral channels).

Spatial Data Reduction. The training dataset consisted of 72 hyperspectral images (24 field pea samples with three replications). A customized software package was used to convert each hyperspectral image file into text format and to compile them into a single file. The hyperspectral imaging data were spatially binned into five resolutions: (1) 50.000 pixels/mm² (1 × 1 pixel grid), (2) 12.500 pixels/mm² (2 × 2 pixel grid), (3) 2.000 pixels/mm² (5 × 5 pixel grid), (4) 0.500 pixel/mm² (10 × 10 pixel grid), and (5) 0.125 pixel/mm² (20 × 20 pixel grid). Without spatial binning, each field pea consisted of approximately 4000 pixels, so with a spatial binning of 20 × 20 approximately 5–10 pixels were acquired from each field pea (Fig. 3). We considered it inappropriate to conduct spatial binning beyond the 20 × 20 level because that would have created files with very low numbers of pixels per field pea.

The spatial binning of 20 × 20 resulted in each hyperspectral image cube consisting of approximately 160 pixels (5–10 pixels from each individual field pea). To classify the same number of pixels in all the training datasets, we conducted the spatial and spectral binnings and randomly selected 160 pixels from each hyperspectral image. Hence, 11 520 pixels (160 pixels × 72 hyperspectral images) were classified in each of the training datasets.

Spectral Data Reduction. In the 401–879 nm range, we obtained spectral data reduction from binning

operations into eight different channel resolutions:¹⁴ 230 spectral channels (spectral resolution of 2.1 nm, binning 1 × 1), 115 spectral channels (spectral resolution of 4.2 nm, binning 1 × 2), 77 spectral channels (spectral resolution of 6.3 nm, binning 1 × 3), 57 spectral channels (spectral resolution of 8.4 nm, binning 1 × 4), 46 spectral channels (spectral resolution of 10.5 nm, binning 1 × 5), 23 spectral channels (spectral resolution of 21.0 nm, binning 1 × 10), 10 spectral channels (spectral resolution of 50.4 nm, binning 1 × 24), and 5 spectral channels (spectral resolution of 100.8 nm, binning 1 × 48).

Data Analysis. We conducted all data processing and analysis in R software.³⁸ To exclude reflection data representing shadows and stochastic noise, we applied a radiometric filter, so that only reflection profiles (pixels) were included, when reflection values in three spectral channels at 432 nm (R432), 481 nm (R481), and 800 nm (R800) met the following criteria:

$$R432 < 0.15 \quad \text{and} \quad R481 < 0.35 \quad \text{and} \quad R800 > 0.30$$

This radiometric filter was deployed to address concerns about the adverse effects of projection angles when high-resolution hyperspectral imaging data are acquired from round objects, such as seeds.³⁷ Deploying this radiometric filter was intended to decrease the within-class variance and therefore to improve the likelihood of successful between-class separation. For each of the 40 datasets (five spatial resolutions times eight spectral resolutions), we used forward stepwise LDA with forward selection of spectral channels²⁴ to select the spectral channels with more than a 0.01% contribution to the classification of reflection profiles based on a dummy variable (pea weevil infested, INF = 1; non-infested, INF = 0). The spectral channels selected from the forward stepwise LDA were used to

TABLE I. Classification accuracies of field peas with different spatial and spectral resolutions.

Spatial resolution (pixels/mm ²)	Spectral resolution (nm)	Number of spectral channels	Number of spectral channels selected by LDA	LDA			SVM		
				Non-infested (%)	Infested (%)	Overall (%)	Non-infested (%)	Infested (%)	Overall (%)
50	2.1	230	35	77.1	82.1	79.6	83.5	83.0	83.3
	4.2	115	26	77.8	81.7	79.7	82.8	83.4	83.2
	6.3	77	23	78.5	81.6	80.1	83.0	82.8	83.0
	8.4	57	20	77.8	82.1	80.0	82.9	83.1	83.0
	10.5	46	14	77.9	82.1	80.0	82.7	82.9	82.8
	21	23	7	78.5	81.5	80.0	82.6	82.8	82.7
	50.4	10	4	76.9	81.9	79.4	82.3	82.3	82.3
	100.8	5	4	79.4	81.1	80.3	82.8	82.3	82.6
	2.1	230	25	77.5	83.8	80.6	84.9	82.0	83.3
	4.2	115	17	77.9	83.5	80.7	85.1	81.5	83.1
12,500	6.3	77	18	77.6	83.0	80.3	84.9	82.1	83.4
	8.4	57	15	77.7	83.3	80.5	85.1	81.8	83.3
	10.5	46	15	78.2	83.4	80.8	85.2	81.9	83.4
	21	23	8	79.3	83.0	81.1	84.9	82.1	83.4
	50.4	10	6	79.4	83.0	81.2	83.7	81.3	82.4
	100.8	5	4	79.2	83.6	81.3	83.8	81.4	82.5
	2.1	230	22	79.2	83.4	81.3	85.2	83.7	84.4
	4.2	115	18	78.6	83.8	81.2	85.6	83.5	84.5
	6.3	77	16	78.8	83.9	81.3	85.7	83.2	84.4
	8.4	57	14	78.7	83.8	81.3	85.8	83.7	84.7
2,000	10.5	46	15	78.7	83.9	81.3	86.1	83.7	84.8
	21	23	8	79.0	83.5	81.3	85.2	83.0	84.0
	50.4	10	6	78.3	83.2	80.7	84.7	82.8	83.7
	100.8	5	4	79.3	83.6	81.5	83.0	82.6	82.8
	2.1	230	29	80.4	85.0	82.7	87.3	84.9	85.9
	4.2	115	20	80.4	84.3	82.4	86.9	83.7	85.1
	6.3	77	23	80.3	84.8	82.6	87.2	84.2	85.5
	8.4	57	20	79.7	85.6	82.6	87.5	84.3	85.8
	10.5	46	18	79.8	85.4	82.6	86.9	84.2	85.4
	21	23	9	80.6	84.2	82.4	86.4	84.8	85.5
0.500	50.4	10	5	80.7	84.8	82.7	85.2	83.3	84.2
	100.8	5	4	80.5	84.8	82.7	85.0	82.8	83.8
	2.1	230	43	82.5	82.9	83.3	85.5	85.5	85.5
	4.2	115	31	82.0	84.6	83.3	86.0	85.1	85.5
	6.3	77	26	81.6	83.9	82.7	85.8	85.2	85.5
	8.4	57	24	81.2	83.6	82.4	85.5	84.5	85.0
	10.5	46	19	81.3	83.2	82.2	85.2	86.0	85.6
	21	23	12	81.6	83.3	82.4	84.8	85.6	85.2
	50.4	10	4	82.7	80.8	81.8	83.0	84.3	83.7
	100.8	5	4	84.4	81.4	81.4	83.1	83.7	83.4
0.125	2.1	230	43	82.5	82.9	83.3	85.5	85.5	85.5
	4.2	115	31	82.0	84.6	83.3	86.0	85.1	85.5
	6.3	77	26	81.6	83.9	82.7	85.8	85.2	85.5
	8.4	57	24	81.2	83.6	82.4	85.5	84.5	85.0
	10.5	46	19	81.3	83.2	82.2	85.2	86.0	85.6
	21	23	12	81.6	83.3	82.4	84.8	85.6	85.2
	50.4	10	4	82.7	80.8	81.8	83.0	84.3	83.7
	100.8	5	4	84.4	81.4	81.4	83.1	83.7	83.4

develop SVM models.²² The Gaussian kernel function was used to solve the nonlinear classification of the samples in a high-dimensional feature space:

$$k(x_1, x_2) = \exp\left(-\gamma \|(x_1 - x_2)\|^2\right) \quad (1)$$

where $x_1 - x_2$ denotes the difference between two pixels and γ (1/[data dimension or number of spectral channels]) determines the width and tunes the smoothing of the SVM function.¹² In addition, in SVM models a separate parameter C is a penalty factor to control for error, and it was set to 1.

After developing the LDA and SVM classification models based on the training datasets, we used each model to classify the 30 independent validation samples having the same spatial and spectral resolutions. A paired t test was conducted to compare the classification accuracies of the LDA models and SVM models. We used an ANOVA to examine the effects of spatial and spectral resolutions on the classification accuracies. In the ANOVA, means were compared with least square difference (LSD) ($P = 0.05$).

RESULTS

Spectral Data Reduction. On the basis of stepwise LDA, the spectral channels that contributed the most to the classification of infested and non-infested field pea samples were located in the following portions of the examined spectrum: 486–516 nm, 723–744 nm, and 765–806 nm. In all the classifications performed, the stepwise LDA reduced the number of spectral channels included in training models by 20–90% (Table I). The percentage of spectral channels included in training models was negatively associated with the spectral resolution (number of available spectral channels). That is, at the highest spectral resolution of 2.1 nm (230 spectral channels available as explanatory variables), only 14% of the spectral channels were included in training models across the five spatial resolutions. In datasets with the lowest spectral resolution of 100.8 nm (five spectral channels available as explanatory variables), 80% of the spectral channels were included in training models. Thus, the highest level of spectral data reduction was obtained through the use of data acquired at a high spectral resolution. Quite noteworthy, there

FIG. 4. Average accuracies of LDA and SVM classification models across spectral resolutions.

was no clear relationship between spectral resolution and classification accuracy (Table I and Fig. 4). The eight spectral resolutions represented up to a 98% data reduction (using 5 spectral channels compared to 230), but SVMs varied only about 1.5 percentage points and LDAs less than 0.5 percentage point in classification accuracies across these spectral resolutions (Fig. 4). Thus, spectral binning effectively markedly reduced the input data size without compromising classification accuracy.

Spatial Data Reduction. There was generally a negative correlation between pixel resolution and classification accuracy (Fig. 5). The ANOVA showed that there was significant difference among the different pixel resolutions ($P < 0.01$; LSD = 0.6%) in both the LDA and SVM models. In SVM models, 10×10 spatial binning (0.500 pixel/mm²) had the highest classification accuracy (85.2%), but there was no significant difference in classification accuracy between the 10×10 and 20×20 spatial binnings ($P = 0.10$). In LDA models, 10×10 spatial binning (0.125 pixel/mm²) had the highest classification accuracy (82.6%). Similar to SVM models, the classification accuracy of binning 20×20 was not significantly different from the classification accuracy of binning 10×10 ($P = 0.61$). Classification accuracies increased with pixel resolution reduction from 50 pixels/mm² (binning 1×1) to 0.5 pixel/mm² (binning 10×10). When the spatial resolution was reduced further, we found a decrease in classification accuracy. Thus, there appeared to be an optimum pixel resolution around 0.500 pixel/mm².

Overall Performances of the Classifications. The overall LDA and SVM classification results for the different pixel and spectral resolutions are summarized in Table I. A paired t -test showed that SVM had significantly higher overall classification accuracies than LDA ($P < 0.01$). There was no significant difference in classifying pea weevil-infested peas between SVM and LDA ($P = 0.77$). But for non-infested field peas, the classification accuracy of SVM (average 84.9%) was significantly higher than the accuracy of LDA (average

FIG. 5. Average accuracies of LDA and SVM classification models across spatial resolutions.

79.6%) ($P < 0.01$). Thus, the significantly higher classification accuracy obtained with SVM (compared to LDA) was attributed to lower levels of false negatives with SVM.

DISCUSSION

We demonstrated that classification accuracy could be increased through spatial binning (reduction in pixel resolution). We also demonstrated that about a 98% data reduction through spectral binning had a negligible effect on the classification accuracy. The SVM classifications outperformed the LDA classifications across most combinations of spectral and spatial resolutions. The analysis of high-dimensional and large datasets is associated with a range of analytical challenges, including the Hughes phenomenon^{10,22,39} or the principle of parsimony.²³ This challenge is caused by the high number of spectral channels or dimensions used to separate classes in that using a high number of spectral channels as explanatory variables also increases the number of statistical parameters defining the classes. If the number of training samples used to generate the statistical parameters is similar to or less than the number of explanatory variables (channels), then there is a distinct risk of model overfitting, which results in reduced classification accuracy. Mathematical research into the Hughes phenomenon is ongoing,⁴⁰ and the phenomenon is well recognized in classifications of reflection data. Furthermore, the Hughes phenomenon highlights the importance of avoiding an excessive number of explanatory variables, so it is important to reduce the size of the input data through spectral binning as much as possible without compromising classification accuracy. In a recent study by Nansen et al.,⁴¹ reflection data were acquired from two crop-insect pest systems: (1) potato psyllid (*Bactericera cockerelli*) infestations of individual bell pepper plants (*Capsicum annuum*) and (2) sugarcane borer (*Diatraea saccharalis*) infestations of individual maize plants (*Z. mays*). Reflection datasets were analyzed based on seven spectral resolutions, and it was demonstrated that the size of the initial input data

could be reduced fourfold without loss of classification accuracy.

Several studies examined the importance of spectral data reduction in the quality control of food products using hyperspectral imaging analysis. Ariana and Renfu¹⁴ used five pixel resolutions (5, 10, 20, 40, and 60 nm) to classify pickling cucumbers (*Cucumis sativus*) and whole pickles with and without internal defects based on partial least squares discriminant analysis. The authors found the highest classification accuracy at pixel resolutions of 20 and 40 nm (compared to pixel resolutions of 5, 10, and 60 nm). In a study analyzing hyperspectral images from livestock feed materials and ruminant-derived bonemeal,¹⁹ different numbers of channels, ranging from 2 to 64, were selected to develop SVM models. Support vector machine models with 48 selected channels from evolutionary search method were found to have the highest classification accuracy. This also highlights the importance of channel selection and, therefore, underscores the importance of deploying stepwise LDA or evolutionary search methods when developing reflection-based classification models. We are unclear about the discrepancy between the published findings on the importance of spectral resolution and the results in the current study; we found pixel resolution to be of limited importance. The possible reason for the difference is the greatly diverse composition of food and contaminants in the published studies, causing challenges for detecting food contaminants.¹⁹ We also suspect that the marked darkening (and therefore fairly easily distinguishable separation of non-infested from infested field peas) may be partially responsible for this.

Regarding the importance of spatial binning, we are unaware of published studies on the importance of spatial resolution for the analysis of ground-based (high-spatial-resolution) hyperspectral imaging data. Previous research involving airborne imaging (mineral mapping) demonstrated that spatial resolution and classification accuracy are generally positively correlated.¹⁸ The current dataset did not permit effective evaluations of spatial data reductions beyond the 20×20 binning of pixels because the original reflection data were acquired using a spatial resolution of 50 pixels/mm² and each field pea comprised approximately 4000 pixels. Our comparisons of LDA and SVM models showed that SVMs were generally associated with the highest classification accuracies, as in previously published studies. In our study, SVM had a significantly higher classification accuracy than LDA, mainly because LDA was associated with a significantly higher level of false negative error.

In conclusion, we demonstrated that a spectral resolution of 10.5 nm and a 10×10 spatial binning (0.500 pixel/mm²) corresponded to a 99.96% data reduction of the original (unbinned) data and that this optimized data reduction was associated with a SVM-based classification accuracy of 85.8% (87.5% for pea weevil–non-infested field peas and 84.3% for pea weevil–infested field peas). In comparison, the original (unbinned) dataset was classified with a classification accuracy of 83.3%. An increase of 2.5 percentage points in classification accuracy may at first glance appear to be modest or less impressive. However, when the

overall classification accuracy is above 83%, there are fewer than 17 percentage points of error to improve on. Thus, a classification accuracy of 85.8% instead of 83.3% is equivalent to 15% less error (2.5/16.7) and therefore a considerable improvement. Furthermore, this improved classification accuracy was obtained after a 99.96% data reduction, which demonstrates that both processing speed and classification accuracy can be improved simultaneously.

ACKNOWLEDGMENTS

This research was partly supported by ARC-Linkage Grant LP0775351 and by a GRDC (Grain Research and Development Corporation) grant, UWA00144.

1. C. Nansen, X. Zhang, N. Aryamanesh, G. Yan. "Use of Variogram Analysis to Classify Field Peas with and Without Internal Defects Caused by Weevil Infestation". *J. Food Eng.* 2014. 123: 17–22.
2. C.B. Singh, D.S. Jayas, J. Paliwal, N.D.G. White. "Identification of Insect-Damaged Wheat Kernels Using Short-Wave Near-Infrared Hyperspectral and Digital Colour Imaging". *Comput. Electron. Agr.* 2010. 73(2): 118–125. doi:10.1016/j.compag.2010.06.001.
3. C.B. Singh, D.S. Jayas, J. Paliwal, N.D.G. White. "Detection of Insect-Damaged Wheat Kernels Using Near-Infrared Hyperspectral Imaging". *J. Stored Prod. Res.* 2009. 45(3): 151–158. doi: http://dx.doi.org/10.1016/j.jspr.2008.12.002.
4. F. Huang, B.L. Leonard, S.H. Moore, D.R. Cook, J. Baldwin, K.V. Tindall, D.R. Lee. "Allele Frequency of Resistance to *Bacillus thuringiensis* Cry1Ab Corn in Louisiana Populations of Sugarcane Borer (Lepidoptera: Crambidae)". *J. Econ. Entomol.* 2008. 101(2): 492–498.
5. J. Wang, K. Nakano, S. Ohashi. "Non-Destructive Detection of Internal Insect Infestation in Jujubes Using Visible and Near-Infrared Spectroscopy". *Postharvest Biol. Technol.* 2011. 59(3): 272–279. doi:10.1016/j.postharvbio.2010.09.017.
6. J. Wang, K. Nakano, S. Ohashi, K. Takizawa, J.G. He. "Comparison of Different Modes of Visible and Near-Infrared Spectroscopy for Detecting Internal Insect Infestation in Jujubes". *J. Food Eng.* 2010. 101(1): 78–84. doi:10.1016/j.jfoodeng.2010.06.011.
7. C. Nansen. "Use of Variogram Parameters in Analysis of Hyperspectral Imaging Data Acquired from Dual-Stressed Crop Leaves". *Remote Sens.* 2012. 4(1): 180–193. doi:10.3390/rs4010180.
8. C. Nansen, T. Macedo, R. Swanson, D.K. Weaver. "Use of Spatial Structure Analysis of Hyperspectral Data Cubes for Detection of Insect-Induced Stress in Wheat Plants". *Int. J. Remote Sens.* 2009. 30(10): 2447–2464. doi:10.1080/01431160802552710.
9. V. Vapnik. *The Nature of Statistical Learning Theory*. New York: Springer-Verlag, 1999. 2nd ed.
10. H. Lu, H. Zheng, Y. Hu, H. Lou, X. Kong. "Bruise Detection on Red Bayberry (*Myrica rubra* Sieb. & Zucc.) Using Fractal Analysis and Support Vector Machine". *J. Food Eng.* 2011. 104(1): 149–153.
11. T. Rumpf, A.-K. Mahlein, U. Steiner, E.-C. Oerke, H.-W. Dehne, L. Plümer. "Early Detection and Classification of Plant Diseases with Support Vector Machines Based on Hyperspectral Reflectance". *Comput. Electron. Agr.* 2010. 74(1): 91–99.
12. B. Qi, C. Zhao, E. Youn, C. Nansen. "Use of Weighting Algorithms to Improve Traditional Support Vector Machine Based Classifications of Reflectance Data". *Opt. Express.* 2011. 19(27): 26816–26826.
13. L. Jiang, B. Zhu, X. Rao, G. Berney, Y. Tao. "Discrimination of Black Walnut Shell and Pulp in Hyperspectral Fluorescence Imagery Using Gaussian Kernel Function Approach". *J. Food Eng.* 2007. 81(1): 108–117.
14. D.P. Ariana, L. Renfu. "Evaluation of Internal Defect and Surface Color of Whole Pickles Using Hyperspectral Imaging". *J. Food Eng.* 2010. 96(4): 583–590. doi:http://dx.doi.org/10.1016/j.jfoodeng.2009.09.005.
15. A.A. Gowen, F. Marini, C. Esquerre, C. O'Donnell, G. Downey, J. Burger. "Time Series Hyperspectral Chemical Imaging Data: Challenges, Solutions and Applications". *Anal. Chim. Acta.* 2011. 705(1–2): 272–282.
16. T.V. Bandos, L. Bruzzone, G. Camps-Valls. "Classification of Hyperspectral Images with Regularized Linear Discriminant Analysis". *IEEE Trans. Geosci. Remote Sens.* 2009. 47(3): 862–873.

17. G. Camps-Valls, L. Gómez-Chova, J. Calpe-Maravilla, J.D. Martín-Guerrero, E. Soria-Olivas, L. Alonso-Chordá, J. Moreno. "Robust Support Vector Method for Hyperspectral Data Classification and Knowledge Discovery". *IEEE Trans. Geosci. Remote Sens.* 2004. 42(7): 1530-1542.
18. F.A. Kruse, J.V. Taranik, M. Coolbaugh, J. Michaels, E.F. Littlefield, W.M. Calvin, B.A. Martini. "Effect of Reduced Spatial Resolution on Mineral Mapping Using Imaging Spectrometry—Examples Using Hyperspectral Infrared Imager (HyspIRI)-Simulated Data". *Remote Sens.* 2011. 3(8): 1584-1602.
19. P. Wilcox, T.M. Horton, E. Youn, M.K. Jeong, D. Tate, T. Herrman, C. Nansen. "Evolutionary Refinement Approaches for Band Selection of Hyperspectral Images with Applications to Automatic Monitoring of Animal Feed Quality". *Intell. Data Anal.* 2014. 18(1): 25-42.
20. H. Wold. "Estimation of Principal Components and Related Models by Iterative Least Squares". In: P.R. Krishnaiah, editor. *Multivariate Analysis*. New York: Academic Press, 1966. Pp. 391-420.
21. S. De Backer, P. Kempeneers, W. Debruyn, P. Scheunders. "A Band Selection Technique for Spectral Classification". *IEEE Geosci. Remote S.* 2005. 2(3): 319-323.
22. B. Guo, S.R. Gunn, R.I. Damper, J.D.B. Nelson. "Customizing Kernel Functions for SVM-Based Hyperspectral Image Classification". *IEEE Trans. Geosci. Remote Sens.* 2008. 17(4): 622-629.
23. D.M. Hawkins. "The Problem of Overfitting". *J. Chem. Inf. Comput. Sci.* 2004. (44): 1-12. doi:10.1021/ci0342472.
24. R.A. Fisher. "The Use of Multiple Measurements in Taxonomic Problems". *Ann. Hum. Genet.* 1936. 7(2): 179-188.
25. P.B. Garcia-Allende, O.M. Conde, J. Mirapeix, A. Cobo, J.M. Lopez-Higuera. "Quality Control of Industrial Processes by Combining a Hyperspectral Sensor and Fisher's Linear Discriminant Analysis". *Sensor Actuators, B.* 2008. 129(2): 977-984. doi:10.1016/j.snb.2007.09.036.
26. G. ElMasry, N. Wang, C. Vigneault, J. Qiao, A. ElSayed. "Early Detection of Apple Bruises on Different Background Colors Using Hyperspectral Imaging". *LWT—Food Sci. Techn.* 2008. 41(2): 337-345.
27. A.A. Gowen, M. Taghizadeh, C.P. O'Donnell. "Identification of Mushrooms Subjected to Freeze Damage Using Hyperspectral Imaging". *J. Food Eng.* 2009. 93(1): 7-12.
28. C. Nansen, T. Herrman, R. Swanson. "Machine Vision Detection of Bonemeal in Animal Feed Samples". *Appl. Spectrosc.* 2010. 64(6): 637-643.
29. United Nations Food and Agriculture Organization (FAO). *Worldwide Regulations for Mycotoxins in Food and Feed in 2003*. FAO Food and Nutrition Paper 81. Rome: UN FAO, 2004.
30. K. McPhee. "Dry Pea Production and Breeding—a Mini-Review". *Food, Agric. Environ.* 2003. 1(1): 64-69.
31. S.L. Clement, D.C. Hardie, L.R. Elbersen. "Variation Among Accessions of *Pisum fulvum* for Resistance to Pea Weevil". *Crop Sci.* 2002. 42(6): 2167-2173.
32. F. Marzo, A. Aguirre, M.V. Castiella, R. Alonso. "Fertilization Effects of Phosphorus and Sulfur on Chemical Composition of Seeds of *Pisum sativum* L. and Relative Infestation by *Bruchus pisorum* L". *J. Agric. Food Chem.* 1997. 45(5): 1829-1833.
33. S.L. Clement, K.E. McPhee, L.R. Elbersen, M.A. Evans. "Pea Weevil, *Bruchus pisorum* L. (Coleoptera: Bruchidae), Resistance in *Pisum sativum* × *Pisum fulvum* Interspecific Crosses". *Plant Breed.* 2009. 128(5): 478-485. doi:10.1111/j.1439-0523.2008.01603.x.
34. N. Aryamanesh, O. Byrne, D.C. Hardie, T. Khan, K.H.M. Siddique, G. Yan. "Large-Scale Density-Based Screening for Pea Weevil Resistance in Advanced Backcross Lines Derived from Cultivated Field Pea (*Pisum sativum*) and *Pisum fulvum*". *Crop Pasture Sci.* 2012. 63(7): 612-618.
35. O.M. Byrne, D.C. Hardie, T.N. Khan, J. Speijers, G. Yan. "Genetic Analysis of Pod and Seed Resistance to Pea Weevil in a *Pisum sativum* × *P. fulvum* Interspecific Cross". *Aust. J. Agric. Res.* 2008. 59(9): 854-862.
36. C. Nansen, N. Abidi, A.J. Sidumo, A.H. Gharalari. "Using Spatial Structure Analysis of Hyperspectral Imaging Data and Fourier Transformed Infrared Analysis to Determine Bioactivity of Surface Pesticide Treatment". *Remote Sens.* 2010. 2(4): 908-925.
37. C. Nansen, M. Kolomiets, X. Gao. "Considerations Regarding the Use of Hyperspectral Imaging Data in Classifications of Food Products, Exemplified by Analysis of Maize Kernels". *J. Agric. Food Chem.* 2008. 56(9): 2933-2938. doi:10.1021/jf073237o.
38. R Development Core Team, . R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing, 2007.
39. G.F. Hughes. "On the Mean Accuracy of Statistical Pattern Recognizers". *IEEE Trans. Inf. Theory.* 1968. 14(1): 55-63.
40. P. Lavergne, V. Patilea. "Breaking the Curse of Dimensionality in Nonparametric Testing". *J. Econometrics.* 2008. 143(1): 103-122.
41. C. Nansen, L.D. Geremias, Y. Xue, F. Huang, J.R. Parra. "Agricultural Case Studies of Classification Accuracy, Spectral Resolution, and Model Over-Fitting". *Appl. Spectrosc.* 2013. 67(11): 1332-1338. doi:10.1366/12-06933.

Queries for apls-69-04-10

1. The sentence beginning, “Widely used to classify...” was rewritten to avoid starting a sentence with an abbreviation. Please check to ensure we have maintained your meaning here.
2. Please be sure to provide manufacturers or sources of seed and equipment throughout.