Article

Early Detection of *Aspergillus parasiticus* Infection in Maize Kernels Using Near-Infrared Hyperspectral Imaging and Multivariate Data Analysis

Xin Zhao [1], Wei Wang [1,*], Xuan Chu [1], Chunyang Li [2] and Daniel Kimuli [1]

1   College of Engineering, China Agricultural University, Beijing 100083, China; zx1992@cau.edu.cn (X.Z.); cauchx0105@163.com (X.C.); kimuli.daniel@yahoo.com (D.K.)
2   Institute of Food Science and Technology, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China; lichunyang968@126.com
*   Correspondence: playerwxw@cau.edu.cn; Tel.: +86-10-6273-7288

Abstract: Fungi infection in maize kernels is a major concern worldwide due to its toxic metabolites such as mycotoxins, thus it is necessary to develop appropriate techniques for early detection of fungi infection in maize kernels. Thirty-six sterilised maize kernels were inoculated each day with *Aspergillus parasiticus* from one to seven days, and then seven groups (D_1, D_2, D_3, D_4, D_5, D_6, D_7) were determined based on the incubated time. Another 36 sterilised kernels without inoculation with fungi were taken as control (DC). Hyperspectral images of all kernels were acquired within spectral range of 921–2529 nm. Background, labels and bad pixels were removed using principal component analysis (PCA) and masking. Separability computation for discrimination of fungal contamination levels indicated that the model based on the data of the germ region of individual kernels performed more effectively than on that of the whole kernels. Moreover, samples with a two-day interval were separable. Thus, four groups, DC, D_{1-2} (the group consisted of D_1 and D_2), D_{3-4} (D_3 and D_4), and D_{5-7} (D_5, D_6, and D_7), were defined for subsequent classification. Two separate sample sets were prepared to verify the influence on a classification model caused by germ orientation, that is, germ up and the mixture of germ up and down with 1:1. Two smooth preprocessing methods (Savitzky-Golay smoothing, moving average smoothing) and three scatter-correction methods (normalization, standard normal variate, and multiple scatter correction) were compared, according to the performance of the classification model built by support vector machines (SVM). The best model for kernels with germ up showed the promising results with accuracies of 97.92% and 91.67% for calibration and validation data set, respectively, while accuracies of the best model for samples of the mixed kernels were 95.83% and 84.38%. Moreover, five wavelengths (1145, 1408, 1935, 2103, and 2383 nm) were selected as the key wavelengths in the discrimination of fungal contamination levels. In general, near-infrared hyperspectral imaging can be used for early detection of fungal contamination in maize kernels.

Keywords: hyperspectral imaging; maize kernel; *Aspergillus parasiticus*; early detection

1. Introduction

Maize (*Zea mays* L.) is one main cereal crop that has been broadly cultivated worldwide for human consumption and animal feed. However, maize is susceptible to be infected with toxigenic fungi, such as *Fusarium* spp. (*F. verticillioides*, *Gibberella moniliformis*, *F. graminearum*) and *Aspergillus* spp. (*A. niger*, *A. flavus*, *A. parasiticus*), in the field and/or post-harvest conditions. Different toxigenic fungi can produce relevant toxic metabolites, for example, aflatoxins by *A. flavus* and *A. parasiticus*,

ochratoxins by *A. niger*, fumonisins by *Fusarium* spp. [1,2]. Among the various toxins, aflatoxin is highly toxic and carcinogenic to animals and humans [3,4]. Therefore, there has been a mass of studies focusing on detection of aflatoxins and fungi on maize to prevent health threats to people resulting from consumption of infected kernels.

Traditional methods to detect fungal contamination on cereals are microbiological methods and diagnostic media for toxigenic fungi identification, and immunological methods for toxin detection. However, these methods are usually time-consuming, professional, expensive, labour-intensive and contaminative by introducing harmful chemical reagents [2,5–8].

In recent years, hyperspectral imaging, which was originally developed for remote sensing applications, emerges and becomes one of the most potential techniques for fast, non-destructive detection of food quality and safety [9–11]. Hyperspectral imaging integrates conventional imaging and spectroscopy to obtain both spatial and spectral information of the object [12,13]. It allows interactive analysis between spectral and surface properties of samples based on each pixel point of images acquired with a wide wavelength range of spectral such as UV-visible, near infrared (NIR) or infrared (IR) [2]. Data of hyperspectral images are three-dimensional arrays of the form, X ($m \times k \times \lambda$), where m and k are the two axes of spatial information of the image, and λ is the number of wavelengths of the spectra.

Hyperspectral imaging also has become a promising non-destructive technique for detection of toxin and fungi contamination on cereals. Yao [14] studied hyperspectral Bright Green-Yellow Fluorescence (BGYF) imaging to detect single corn kernels contaminated with aflatoxin. Firrao et al. [15] used multispectral imaging (720–940 nm) to predict fumonisin content of milled maize. Wang et al. [16–18] studied the use of hyperspectral imaging, with the range of 400–1000 nm and 1000–2500 nm, to detect aflatoxin B1 on maize kernels, which were artificially titrated on the kernel surface in the laboratory and inoculated with *Aspergillus flavus* conidia in the field, respectively. Kandpal et al. [19] verified the feasibility of short wave infrared (SWIR) hyperspectroscopy to detect aflatoxin on the surface of three different corn varieties (yellow, white, and purple). The results showed that the hyperspectral imaging technique can be used to detect the aflatoxin on or in maize kernels. However, research also found that it is hard to directly detect mycotoxins non-destructively. Therefore, some studies turn around to detect the contamination caused by fungi or mycotoxins, which affects other chemical and optical properties of whole kernels, and these changes can be detected with NIR hyperspectral imaging or spectroscopy [17,20–22]. Moreover, compared to toxin detection, detection of toxigenic fungi on cereals is important for the objective of early detection, which is significant in that it can prevent contamination in the form of mycotoxins in the food chain. Singh et al. and Zhang et al. [23,24] studied the application of near-infrared reflectance hyperspectral imaging for detection of wheat kernels infected with different fungi. Bauriegel et al. [25] studied wheat plants using hyperspectral imaging (400–1000 nm) for early detection of *Fusarium* infection, and results demonstrated that *Fusarium* infestation could not be detected by spectral analysis immediately after infection. Williams et al. [26] used hyperspectral imaging with a spectral range of 1000–2498 nm to study maize kernels inoculated with *F. verticillioides* at the incubated process from 0 to 90 h, and results showed that principal component analysis (PCA) models without pre-processing could discriminate between the images of the control, 17 h after inoculation and 20–90 h, and infection that changes the starch structure or content was noted. Del Fiore et al. [2] studied early detection of maize kernels inoculated with different toxigenic fungi using visible-near infrared (Vis/NIR) hyperspectral imaging (400–1000 nm), and results indicated that the analytical method they presented allowed early detection of fungal contaminants on the maize, starting at 48 h from inoculation and incubation at 30 °C for *A. flavus* and *A. niger*. All of the studies showed the possibility of the application of hyperspectral imaging in fungi detection on cereals. Although some scientific studies were successfully carried out to detect toxin and fungi contamination on maize at an early stage [2,19], studies on detection of fungi contamination on maize are limited, and are applied with different wavelength bands of hyperspectral imaging technologies, such as fluorescence (ultraviolet, UV), 400–1000 nm

(Vis/NIR), and 1000–2500 nm (NIR), for different fungi infection detection, such as *Fusaium* spp. and *Aspergillus* spp. [2,26,27]. Therefore, there are limited studies on detection of *Aspergillus parasiticus* infection in maize kernels using NIR hyperspectral imaging.

The objectives of this work were to study early detection of maize kernels infected with *Aspergillus parasiticus* using NIR hyperspectral imaging. To that end, the specific objectives were to: (1) detect fungal contamination of maize kernels before the onset of visual symptoms; to (2) discriminate contamination and/or damage levels; and to (3) determine optimal wavelengths for subsequent instrument development.

2. Materials and Methods

2.1. Maize Kernel Preparation

Maize kernels (Jingke 968) harvested from Hefei City, Anhui Province, China in 2015 were used for the experiment. Firstly, all kernels were immersed in distilled water for 24 h. Then, in order to remove the natural fungi on both the surface and inside of the kernel, samples were sterilised by rinsing in 75% ethanol and 1% NaClO solution separately. After that, kernels were washed with sterile distilled water and then left to dry in a laminar flow for 2 h. A total of 288 kernels were randomly selected, 36 of which were taken and used as control (DC for short, means without inoculation with fungi).

2.2. Fungal Inoculation

Aspergillus parasiticus (3.6155) strain was obtained from China General Microbiological Culture Collection Center (CGMCC, Beijing, China). The strain was firstly transferred onto potato dextrose agar (Baiaolaibo, Beijing, China) and incubated at 28 °C for 7 days. Spores were obtained from the agar surface by gently rubbing the colonies using a sterile loop. The spore suspension was shaken and then poured through two layers of sterile cheesecloth to remove mycelium. Thereafter, the inoculums concentration (spores/mL) was adjusted to 10^6 spores·mL^{-1} by microscopic enumeration with a haemocytometer (Qiujing, Shanghai, China). Thirty-six sterilised maize kernels were inoculated each day with *Aspergillus parasiticus* from 1 to 7 days. Kernels were inoculated by dipping into the spore suspension for 30 s and were then incubated at 30 °C. The contaminated kernels were divided into seven groups based on the incubated time, that is, incubated for 1 to 7 days, referred to as D_1, D_2, D_3, D_4, D_5, D_6, D_7, respectively. Thirty-six kernels of each group were separately placed onto two Petri dishes (18 kernels per Petri dish). On one Petri dish, the kernels were orientated with germ up. On another one, kernels were germ down. After collecting the images, two kernels were taken away from each group to verify the presence of the fungus based on common fungal isolation and culture methods. The procedure is as follows: firstly, kernels were washed with sterile distilled water, and then the used water was dropped onto rose Bengal agar and incubated at 30 °C for three days.

2.3. NIR Hyperspectral System and Image Calibration

Hyperspectral images were collected by spectral camera (Specim, Spectral Imaging Ltd., Oulu, Finland). The camera was equipped with a cryogenically cooled Mercury-Cadmium-Telluride (MCT) detector. Individual images consisted of 288 spectral bands with the range from 921 to 2529 nm at 5.6 nm resolution. The matched Lumo software (Specim, Spectral Imaging Ltd., Oulu, Finland) was for data acquisition and camera parameters setting. Eight images (DC, D_1, D_2, D_3, D_4, D_5, D_6, D_7) were collected at identical times (the eighth day from the first inoculation, and control group DC samples were prepared at the seventh day) after all inoculations from 1 to 7 days. Each image consisted of the above-mentioned two uncovered Petri dishes. To reduce the influence of illumination and dark current of the charge-coupled-device (CCD) detectors, white and dark references were captured prior to sample images for image correction and calibration. A white reference was acquired based on a standard Teflon white board, and a dark reference was obtained by closing the shutter. The calibrated images were calculated according to Equation (1):

$$R = \frac{R_0 - D}{W - D},\tag{1}$$

where R = calibrated images; R_0 = raw images; D = dark reference; W = white reference.

2.4. Data Analysis

Hyperspectral images were analyzed using ENVI v.5.1 (Exelis Visual Information Solutions Inc., Redlands, CA, USA). Before analysis, spectra range was reduced to 1000–2500 nm for noise reduction. Regions of interest (ROIs) of maize kernels of each image were identified and selected in order to mask the background, Petri dishes and labels in the image. Principal component analysis (PCA) as an unsupervised classification algorithm has the ability to separate data with distinct differences. Thus, PCA was applied on the image for the selection of ROIs of maize kernels [18]. After the selection, inverse PCA was used to reduce noise by removing some PCs that represented noise information, and transform PCs back to wavelengths for subsequent analysis. Then, a mask built with ROIs was applied to obtain a clean image without irrelevant information, such as label and background.

After that, each clean image was resized to remove the blank edges around the kernels, and then all eight of the resized images were integrated to construct a mosaic. It is well-known that PCA can reduce data dimensionality by transforming the raw large number of inter-correlated variables into fewer and orthogonal variables, which are referred to as PCs. PCs are calculated based on the maximum variance of the data, and they are in descending order according to the percentage of variance. Generally, the last PCs are for noise and can be removed. Based on hyperspectral image data, PC loading line plots and PC score images can be displayed and discussed interactively to explore spectra and image information simultaneously. In this research, PCA was applied on the mosaic to reduce data dimensionality and noise to benefit the next classification of different fungal contamination levels.

Computation of ROI separability in ENVI software is usually applied to have a preliminary observation on classification effect of data before application of a classification modeling algorithm. The separation results are calculated on every two classes, represent the separable ability between two classes, and are shown in a list of values for all of the pairs of classes. The separability values vary from 0 to 2. The closer the value to 2, the better the separation for the two classes performed. In this work, the separability for the eight groups was calculated based on two types of data sets, one of which is data of the whole kernel, and the other is data of the germ region. The sample kernels of one Petri dish, on which kernels were orientated with germ up, were used. Elliptical ROIs of the germ region on each kernel were hand-digitized using ENVI software.

The mean reflectance from each ROI of the germ region of kernel samples of the two Petri dishes was calculated. Samples from both of the two Petri dishes made a data set with 1:1 for the number of kernels orientated with germ up and germ down, which was used for discussion on the influence on classification models caused by germ orientation. For the samples orientated with germ down, the germ region was defined as the area symmetrical to the germ portion. Spectral data was analyzed using MATLAB software (The Math Works, Natick, MA, USA).

The raw spectral data (from ROIs) contained scattering noise generated by a camera, random noise, baseline offset, non-uniformity and surface scattering in samples [28]. Therefore, to highlight the effective information, pretreatment techniques were applied. In this study, the mathematical pretreatment techniques used were Savitzky-Golay smoothing (SGS), moving average smoothing (MAS), normalization (NMZ), standard normal variate (SNV), and multiple scatter correction (MSC), as well as their combinations. SGS and MAS are moving window methods used to eliminate overlapping peaks due to random noise [29,30]. NMZ, SNV and MSC were scatter-correction methods and designed to remove the slope variation from the spectra generated by the scatter and variation of particle size and adjust baseline shifts between samples [31–33]. PCA can reduce the dimensionality of data by contracting data into a few descriptive dimensions, denoted principal components,

which represent main variation in the data [34,35]. It is also commonly used as an exploratory tool for data analysis due to the ability of each PC to be displayed graphically [36]. Kennard-Stone (KS) used in this study is a technique for calibration and validation subset partitioning. It is based on maximizing the Euclidean distances between spectral data of samples to cover the multidimensional space [37]. Support vector machines (SVMs) were used for classification modeling in the study. SVMs are proven to present good generalization performance and to be able to model complex nonlinear boundaries through the use of adapted kernel functions [38]. In this work, radial basis function was used and particle swarm optimization (PSO) based on fivefold cross-validation was adopted for optimization of the parameters in SVM [39–41].

3. Results and Discussion

3.1. Hyperspectral Image Analysis

3.1.1. ROIs Extraction Based on PCA and Masking

Processing on image data of D_7 was shown as an example. After PCA was applied, the first three PCs had a total variance percentage of 99.57%, while individual variance percentages were 92.31%, 6.34%, and 0.92% for PC_1, PC_2 and PC_3, respectively. Comparing the score images of the first PCs, although PC_1 had a maximum value of variance percentage, maize kernels seemed to be more distinct from the background on PC_2 and PC_3 score images. Therefore, data of PC_2 and PC_3 were used to make a scatter plot (Figure 1a) to extract object of maize kernels. Five classes were selected based on density slice imagery (Figure 1b), the corresponding pixels of each class were shown in a score image (Figure 1c) where the red class represented the whiteboard, the magenta class the background, the purple class the label, sea green for most parts of the Petri dishes, and yellow for kernels and lesser parts of the Petri dishes. Some parts of the labels produced specular reflection. Hence, they were gathered with the whiteboard in the scatter plot. Since kernels and parts of the Petri dishes were gathered together, it was difficult to separate them clearly in this step. Then, the yellow class was exported to ROI tools and used to make a mask to remove other classes' information. After the mask was applied, retained data also contained parts of the Petri dishes. A scatter plot of PC_3 vs. PC_4 (Figure 2a) was drawn based on the masked data. As shown in Figure 2a, two distinct clusters were apparent. The green cluster was for the parts of Petri dishes (Figure 2b), and exported to make an inverse mask to remove the parts of Petri dishes. After that, a clean image (Figure 2c) with only maize kernels was obtained. Image data of the other treatments (D_1, D_2, D_3, D_4, D_5, D_6, and DC) were processed with the same process steps.
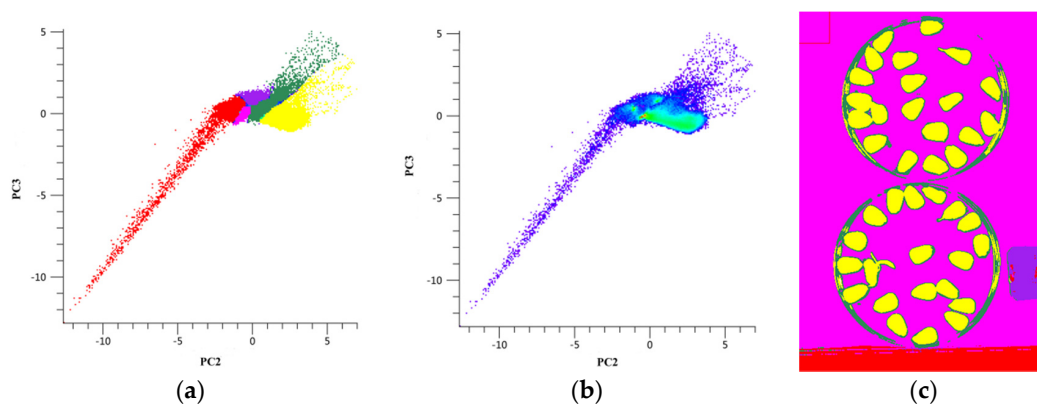

Figure 1. (**a**) PCA scatter plot of PC_2 vs. PC_3 (6.34% and 0.92%) of the image of D_7; (**b**) density scatter plot of PC_2 vs. PC_3; and (**c**) corresponding score image where the red class represented the whiteboard and lighting sections of labels, the magenta represented the background, the purple the label, sea green for most parts of the Petri dishes, and yellow for kernels and lesser parts of the Petri dishes.
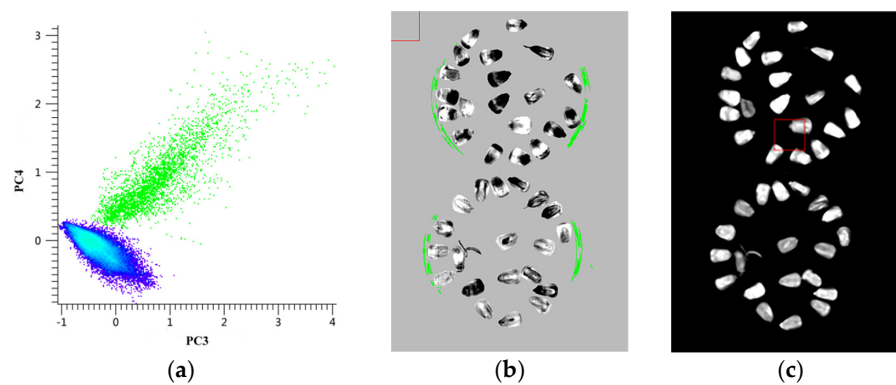
Figure 2. (**a**) PCA scatter plot of PC_3 vs. PC_4 (0.92% and 0.2%) of the masked image of D_7; (**b**) corresponding score image where the green class represented the Petri dishes; and (**c**) a clean image.

3.1.2. Further Processing of Bad Pixels and Separability Computation of Contamination Levels

The mosaic image containing the eight clean images was further resized to remove the blank edges around the kernels. Then, PCA was applied on the mosaic. The first four PCs had a total variance percentage of 99.90%, while PC_5 to PC_8 had a total variance percentage of 0.08%. However, as shown in Figure 3, loading line plot of PC_5 to PC_8 had several sharp peaks. Corresponding wavelengths at the locations of the sharp peaks were 1016, 1067, 1100, 1111, 1117, 1229, and 1263 nm. Correspondingly, images at these monochromatic wavelengths were found to have several bad pixel lines, which images at other wavelengths did not have. For an example, images at 1111 and 1240 nm were shown in Figure 4. Therefore, the sharp peaks in the loading line plot of PC_5 to PC_8 were produced due to noise induced by bad pixel lines. Thus, before subsequent analysis, it is necessary to cancel the data at these wavelengths.

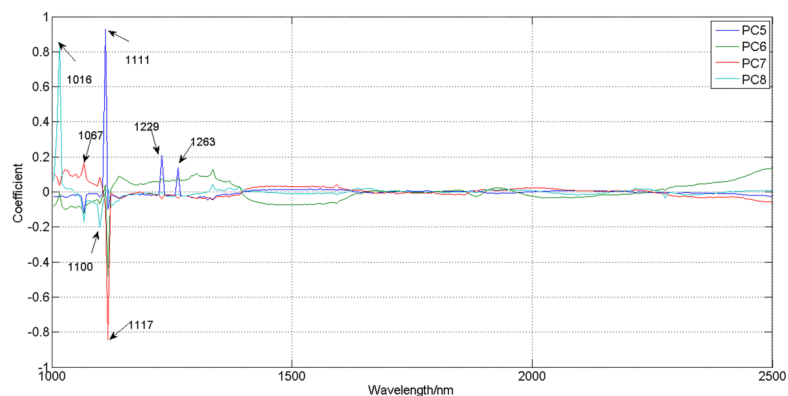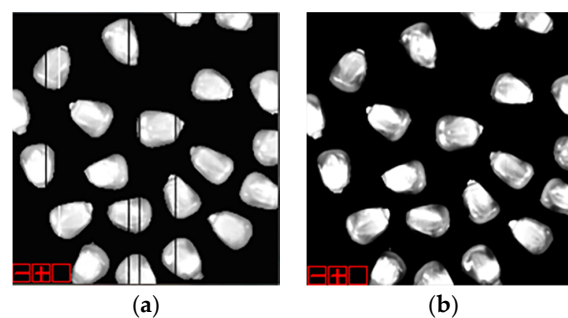**Figure 3.** Loading line plot of PC_5–PC_8 of the mosaic image.

Figure 4. Images at (**a**) 1111 and (**b**) 1240 nm.

In our experiment, only a few kernels of D_6 and D_7 had obvious hyphae on the surface, and almost all of these kernels showed symptomatic appearance of fungal infection originally in the germ region, which is consistent with life experience. Pseudo-color images composited by 1095, 1291, 1543 nm of kernels with symptomatic appearance of fungal infection (have obvious hyphae on kernel surface) and asymptomatic appearance were shown in Figure 5. It was proposed that the germ region data might be better than the whole kernel data for further classification of contamination levels and feature extraction of fungi infection.
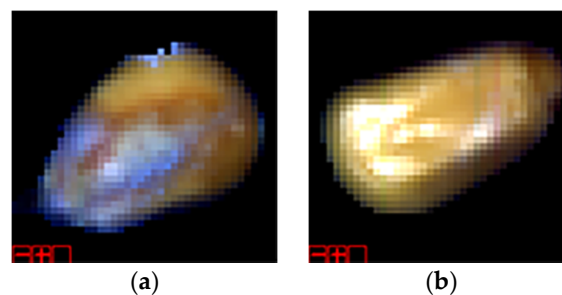

(**a**)   (**b**)

Figure 5. Pseudo-color images of kernels with (**a**) symptomatic appearance and (**b**) asymptomatic appearance composited by 1095, 1291, 1543 nm.

Separability computation in ENVI software was applied to investigate classification effects of whole kernel data and germ region data. The results for 28 (C_8^2) pairs of comparison among eight treatments were plotted as distribution histograms (Figure 6). Compared to the spectral data based on whole kernel data (Figure 6a), the separation values calculated based on germ region data (Figure 6b) were more concentrated at the value of 2. It indicated that data of germ regions were more effective for classification, which might be due to the fact that a greater proportion of fungi start growing from and gather more in germ regions than in whole kernels. Therefore, spectral data of germ regions of all kernel samples were firstly extracted and used for later spectral analysis.
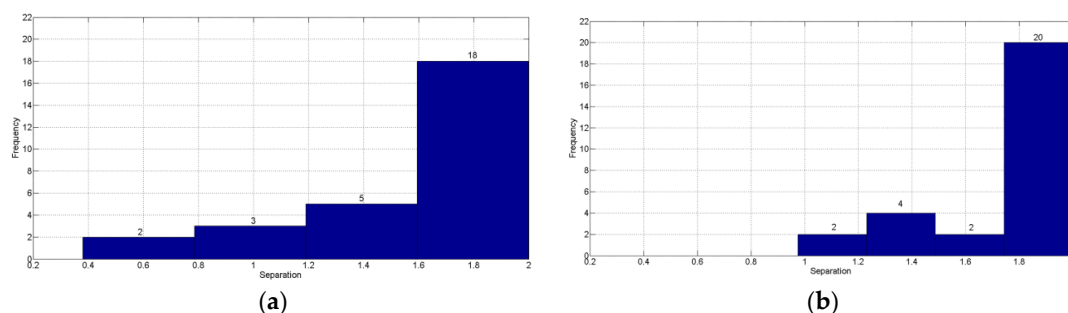

(**a**)   (**b**)

Figure 6. Distribution plots of separability values calculated based on (**a**) whole kernel data and (**b**) germ region data.

3.2. Spectral Analysis

Average spectral curves of the germ region of control and infected kernels with different days were shown in Figure 7. According to the separability computation, some separability values were less than 1.4, which meant that those groups were not easily separable from each other. Hence, four new groups were defined, that is, DC, D_{1-2}, D_{3-4}, D_{5-7} (D_{1-2} group consisted of D_1 and D_2; D_{3-4} group consisted of D_3 and D_4; D_{5-7} group consisted of D_5, D_6 and D_7.). Influence on the performance of the classification model caused by germ orientation was further investigated. For that purpose, two sets of data used for modeling were extracted. One set was the ROI data of kernels orientated with germ up, while the other was the ROI data of the mixture of kernels orientated with germ up and down with 1:1.
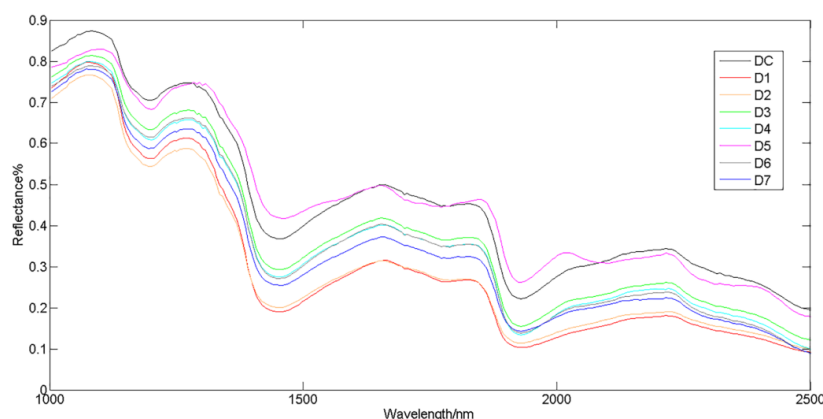
Figure 7. Average spectra of the germ region of all kernels under 8 different treatments.

3.2.1. Classification Based on Data of Maize Kernels with Germ Up

Samples of the Petri dish with kernels orientated with germ up were used, a total of 144 samples made up the data set, with 18 samples of each of the eight treatments. Ninety-six of 144 kernels were selected by using the Kennard-Stone (KS) algorithm as the calibration set, while the other 48 kernels were used as the validation set. Two smooth methods, SGS and MAS, were used firstly to reduce random noise of raw data. In addition, to get rid of the scattering effects caused by different surface roughness and kernel shape, three scatter correction methods, NMZ, SNV and MSC, were applied individually. Discrimination performance of different pretreatment methods and corresponding combinations were shown in Table 1. The performance of SGS and MAS was similar, whether or not they combined with other methods. Results of NMZ were the worst and accuracies of validation (72.92%) were much lower than that of calibration (100%). This might be caused by introducing only the maximum and minimum values of the data set in the algorithm. MAS combined with SNV had the highest classified accuracies of 97.92% and 91.67% for calibration and validation data sets, respectively. The corresponding confusion matrix with characteristics (i.e., sensitivity, specificity, and overall accuracy) was shown in Table 2. Generally, the results were promising with the precision, sensitivity and specificity substantially above 80%, except for low sensitivity (57.14%) for DC in validation, which might be caused by less total sample numbers of the DC group and a large proportion of the number of validation sets compared with the number of calibration sets for the DC group. Moreover, the D_{1-2} group showed the highest separability, which was illogical. The reason may be related to water content in maize sample kernels. In the experiment, kernels were immersed in water to create a humid environment for mold growth and dipped into the spore suspension for inoculation. A portion of water in kernels evaporated rapidly during the initial period after inoculation, which caused a relatively larger difference in water content between the D_{1-2} group and other groups (the DC group was not dipped into the spore suspension). The influence of water content could be verified in Figure 8 (loading line plot for PC_4–PC_6), where several predominant peaks and valleys at 1935 nm (associated with water) were shown. Misclassification might be caused by the size, texture and shape of germ of the samples.

Table 1. Comparison of classification models with different pretreatment methods based on kernel samples orientated with germ up.

Pretreatment Methods	MAS	SGS	MAS + NMZ	SGS + NMZ	MAS + SNV	SGS + SNV	MAS + MSC	SGS + MSC
PCs	8	7	9	9	**11**	9	6	6
Calibration (%)	89.58	88.54	100	100	**97.92**	97.92	82.29	84.38
Validation (%)	87.50	87.50	72.92	72.92	**91.67**	87.50	81.25	83.33

MAS: moving average smoothing (width = 7); SGS: Savitzky-Golay smoothing (width = 7); NMZ: normalization; SNV: standard normal variate; MSC: multiple scatter correction.

Table 2. Confusion matrix of calibration and validation set of the best model based on kernels orientated with germ up.

Actual Result	Predicting Result							
	Calibration				Validation			
	DC	D_{1-2}	D_{3-4}	D_{5-7}	DC	D_{1-2}	D_{3-4}	D_{5-7}
DC	9	0	2	0	4	0	3	0
D_{1-2}	0	22	0	0	0	14	0	0
D_{3-4}	0	0	21	0	0	0	14	1
D_{5-7}	0	0	0	42	0	0	0	12
Precision (%)	100.00	100.00	91.30	100.00	100.00	100.00	82.35	92.31
Sensitivity (%)	81.82	100.00	100.00	100.00	57.14	100.00	93.33	100.00
Specificity (%)	100.00	100.00	97.33	100.00	100.00	100.00	90.91	97.22
Overall accuracy (%)	97.92				91.67			

D_{1-2}: the group consisted of D_1 and D_2; D_{3-4}: the group consisted of D_3 and D_4; D_{5-7}: the group consisted of D_5, D_6 and D_7.

3.2.2. Classification Based on Data of the Mixture of Maize Kernels Orientated Germ Up and Down

For practical application, it was expected that performance for classification was similar regardless of the orientation of kernels. Data of kernels of the two Petri dishes were analyzed and discussed. A total of 288 samples made up the data set, with 36 from each of the eight treatments. In addition, 192 of 288 kernels were selected as the calibration set, while the other 96 kernels were used as the validation set. Results of the data analyzed with the same chemometric algorithms were shown in Table 3. Similar to Table 1, NMZ showed the worst performance among the pretreatment methods and MAS combined with SNV had the highest accuracies. Compared with Table 1, with the same pretreatment methods, the results were worse in Table 3 than that in Table 1, which indicated that germ orientation does have an influence on the classification performance. As for maize kernels with different characters on two sides, i.e., the germ and endosperm side, published literature showed oil-rich germs are colonized by aflatoxin-producing fungi before endosperm tissue [42,43], which makes a distribution difference in the depth of fungi. Moreover, major information in reflectance spectra originate from the surface and several millimeters below the surface of the side facing the camera. Hence, data of kernels with germ up performed more effectively than that of germ down for detecting early growth of fungi. Despite this, accuracies of the best results were 95.83% and 84.38% for calibration and validation, respectively. Both were above 84%, which was also significant for application. The corresponding confusion matrix with characteristics (i.e., sensitivity, specificity, and overall accuracy) was shown in Table 4. The results were similar to Table 2, and slightly worse.

Table 3. Comparison of classification models with different pretreatment methods based on the mixture of kernel samples orientated with germ up and down.

Pretreatment Methods	MAS	SGS	MAS + NMZ	SGS + NMZ	MAS + SNV	SGS + SNV	MAS + MSC	SGS + MSC
PCs	8	7	9	9	**6**	5	6	6
Calibration (%)	88.54	84.90	100	100	**95.83**	89.58	80.21	79.17
Validation (%)	79.17	75.00	67.71	70.83	**84.38**	81.25	77.08	76.04

MAS: moving average smoothing (width = 7); SGS: Savitzky-Golay smoothing (width = 7); NMZ: normalization; SNV: standard normal variate; MSC: multiple scatter correction.

Variance percentages of each of the first six PCs used for the best classification model were listed in Table 5. All of them contributed 99.40% of the total variation information. Then, in order to select the characteristic wavelengths, loading line plot of the first six PCs were analysed. According to Manley et al. [44], PC_1 to PC_3 reflected the variation due to the effect of grain topography, such as

kernel curvature, shape and texture, on spectral variation of NIR hyperspectral images. Loading line plot for PC_4 to PC_6 was shown in Figure 8. Five key wavelengths, ascribed to meaningful chemical functional groups, were pointed out. They were 1145, 1408, 1935, 2103, and 2383 nm. The wavelength of 1146 nm ascribed to the 2nd overtone (OT) of aromatic –CH- group, showed the existence of the benzene ring, which can be attributed to some kinds of the few amino acids such as tyrosine [16]. The wavelength of 1405 nm, with an O–H stretch first overtone attributed to an ROH structure, was associated with starch [26,29]. Furthermore, 1940 nm was ascribed to the O–H bend second overtone of water [29]. A wavelength of 2100 nm was associated with carbohydrate [29]. In addition, 2382 nm was ascribed to the 2nd overtone of C–H stretching from aliphatic compounds [16,29]. These indicated that variation in different groups was mainly caused by changes of ingredient contents of maize germ due to the fungal growth.

Table 4. Confusion matrix of calibration and validation set of the best model based on the mixture of kernels orientated with germ up and down.

| | Predicting Result | | | | | | | |
| | Calibration | | | | Validation | | | |
Actual Result	**DC**	**D_{1-2}**	**D_{3-4}**	**D_{5-7}**	**DC**	**D_{1-2}**	**D_{3-4}**	**D_{5-7}**
DC	18	0	2	1	7	0	7	1
D_{1-2}	1	43	0	0	0	28	0	0
D_{3-4}	0	0	39	3	2	0	26	2
D_{5-7}	1	0	0	84	0	0	3	20
Precision (%)	90.00	100.00	95.12	95.45	77.78	100.00	72.22	86.96
Sensitivity (%)	85.71	97.73	92.86	98.82	46.67	100.00	86.67	86.96
Specificity (%)	98.83	100.00	98.67	96.26	97.53	100.00	84.85	91.67
Overall accuracy (%)	95.83				84.38			

D_{1-2}: the group consisted of D_1 and D_2; D_{3-4}: the group consisted of D_3 and D_4; D_{5-7}: the group consisted of D_5, D_6 and D_7.

Table 5. Variance percentages of the first six PCs.

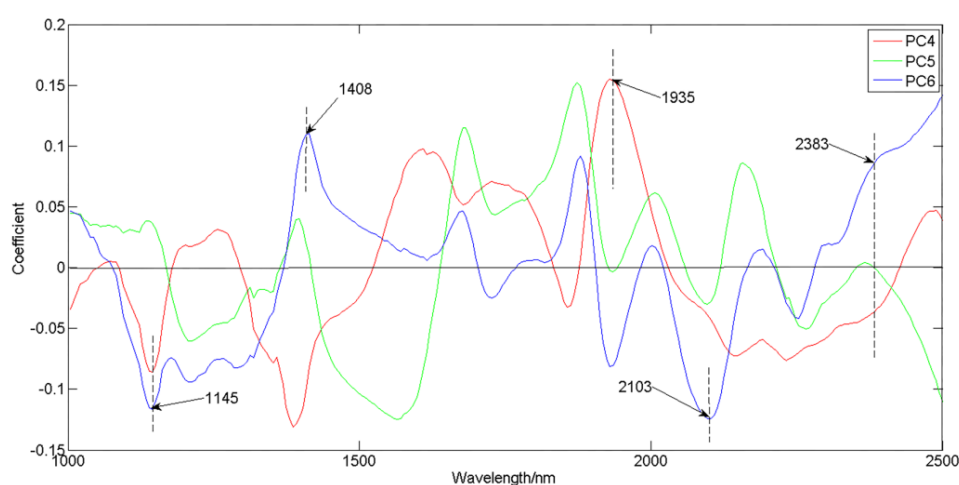| PCs	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6	Total
Variance percent (%)	75.98	16.84	4.37	1.30	0.47	0.44	99.40

Figure 8. Loading line plot for PC_4–PC_6 used in the best classification model.

4. Conclusions

The feasibility of early detection of *Aspergillus parasiticus* infection in maize kernels was verified by using NIR hyperspectral imaging. PCA was applied on image data to extract ROIs from background, labels and others. With loading line plot, wavelengths at which images contained bad pixels were recognized. Germ region data was more effective in discrimination of contamination levels than whole kernel data, which was confirmed by separability computation.

The classification model based on germ region data (of all kernels orientated with germ up) showed the best results with the preprocess method of a combination of MAS and SNV. The accuracies were 97.92% and 91.67% for calibration and validation, respectively, which demonstrated the ability to separate sound kernels and kernels of different contamination levels of $D_{1–2}$ (the group consisted of D_1 and D_2), $D_{3–4}$ (D_3 and D_4), and $D_{5–7}$ (D_5, D_6, and D_7). Germ orientation was found to affect the performance of classified models. Despite this, accuracies of the model, which were built based on data of the mixture of kernels with germ up and down, were 95.83% and 84.38% for calibration and validation respectively, which still showed the possibility for the application of the presented method.

Five key wavelengths (1145, 1408, 1935, 2103, and 2383 nm) were selected according to the loading line plot of PCs used in the model, and were associated with amino acid, starch, water, carbohydrate, and aliphatic compounds, respectively. These are the main ingredients in maize germ, which indicated that variation in different contamination levels was mainly caused by changes of ingredient contents in maize germ due to the fungal growth. The evaporation of water from the maize kernels after inoculation adversely affected the experimental results. Therefore, this process should be noticed in future studies. Furthermore, for future studies, a larger data set of kernels needs to be prepared for further analysis on feature extraction of fungal infection, and different varieties of maize kernels need to be tested to verify the practical application of the presented method.

Author Contributions: W.W. and X.C. conceived and designed the experiments; X.C., X.Z., C.L. and D.K. performed the experiments; X.Z. and X.C. analyzed the data; X.Z. wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Fanelli, C.; Ricelli, A.; Reverberi, M.; Fabbri, A.A.; Pandalai, S.G. Aflatoxins and ochratoxins in cereal grains: An open challenge. *Recent Res. Dev. Crop Sci.* **2004**, *1*, 295–317.
2. Del Fiore, A.; Reverberi, M.; Ricelli, A.; Pinzari, F.; Serranti, S.; Fabbri, A.A.; Bonifazi, G.; Fanelli, C. Early detection of toxigenic fungi on maize by hyperspectral imaging analysis. *Int. J. Food Microbiol.* **2010**, *144*, 64–71. [CrossRef] [PubMed]
3. Teye, E.; Huang, X.; Afoakwa, N. Review on the potential use of near infrared spectroscopy (NIRS) for the measurement of chemical residues in food. *Am. J. Food Sci. Technol.* **2013**, *1*, 1–8.
4. Kelly, J.D.; Eaton, D.L.; Guengerich, F.P.; Coulombe, R.A. Alfatoxin B1 activation in human lung. *Toxicol. Appl. Pharmacol.* **1997**, *144*, 88–95. [CrossRef] [PubMed]
5. Muthomi, J.W.; Ndung'u, J.K.; Gathumbi, J.K.; Mutitu, E.W.; Wagacha, J.M. The occurrence of Fusarium species and mycotoxins in Kenyan wheat. *Crop Prot.* **2008**, *27*, 1215–1219. [CrossRef]
6. Medina-Martínez, M.S.; Martínez, A.J. Mold occurrence and aflatoxin B_1 and fumonisin B_1 determination in corn samples in Venezuela. *J. Agric. Food Chem.* **2000**, *48*, 2833–2836. [CrossRef] [PubMed]
7. Castells, M.; Marín, S.; Sanchis, V.; Ramos, A.J. Distribution of fumonisins and aflatoxins in corn fractions during industrial cornflake processing. *Int. J. Food Microbiol.* **2008**, *123*, 81–87. [CrossRef] [PubMed]
8. Paepens, C.; De Saeger, S.; Sibanda, L.; Barna-Vetro, I.; Léglise, I.; Van Hove, F.; Van Peteghem, C. A flow-through enzyme immunoassay for the screening of fumonisins in maize. *Anal. Chim. Acta* **2004**, *523*, 229–235. [CrossRef]
9. Kim, M.S.; Chen, Y.R.; Mehl, P.M. Hyperspectral reflectance and fluorescence imaging system for food quality and safety. *Trans. ASAE* **2001**, *44*, 721–729.

10. Gowen, A.A.; O'Donnell, C.; Cullen, P.J.; Downey, G.; Frias, J.M. Hyperspectral imaging—An emerging process analytical tool for food quality and safety control. *Trends Food Sci. Technol.* **2008**, *18*, 590–598. [CrossRef]

11. Qin, J.; Chao, K.; Kim, M.S.; Lu, R.; Burks, T.F. Hyperspectral and multispectral imaging for evaluating food safety and quality. *J. Food Eng.* **2013**, *118*, 157–171. [CrossRef]

12. Geladi, P.; Burger, J.; Lestander, T. Hyperspectral imaging: Calibration problems and solutions. *Chemom. Intell. Lab. Syst.* **2004**, *72*, 209–217. [CrossRef]

13. Geladi, P.; Grahn, H.; Burger, J. Multivariate images, hyperspectral imaging: Background and equipment. In *Techniques and Applications of the Hyperspectral Image Analysis*; John Wiley & Sons, Ltd.: West Sussex, UK, 2007; pp. 1–15.

14. Yao, H.; Hruska, Z.; Brown, R.L.; Cleveland, T.E. Hyperspectral bright greenish-yellow fluorescence (BGYF) imaging of aflatoxin contaminated corn kernels. In Proceedings of the Optics for Natural Resources, Agriculture, and Foods, Boston, MA, USA, 1 October 2006.

15. Firrao, G.; Torelli, E.; Gobbi, E.; Raranciuc, S.; Bianchi, G.; Locci, R. Prediction of milled maize fumonisin contamination by multispectral image analysis. *J. Cereal Sci.* **2010**, *52*, 327–330. [CrossRef]

16. Wang, W.; Heitschmidt, G.W.; Ni, X.; Windham, W.R.; Hawkins, S.; Chu, X. Identification of aflatoxin B_1 on maize kernel surfaces using hyperspectral imaging. *Food Control* **2014**, *42*, 78–86. [CrossRef]

17. Wang, W.; Heitschmidt, G.W.; Windham, W.R.; Feldner, P.; Ni, X.; Chu, X. Feasibility of detecting aflatoxin B_1 on inoculated maize kernels surface using Vis/NIR hyperspectral imaging. *J. Food Sci.* **2015**, *80*, M116–M122. [CrossRef] [PubMed]

18. Wang, W.; Lawrence, K.C.; Ni, X.; Yoon, S.C.; Heitschmidt, G.W.; Feldner, P. Near-infrared hyperspectral imaging for detecting Aflatoxin B_1 of maize kernels. *Food Control* **2015**, *51*, 347–355. [CrossRef]

19. Kandpal, L.M.; Lee, S.; Kim, M.S.; Bae, H.; Cho, B.K. Short wave infrared (SWIR) hyperspectral imaging technique for examination of aflatoxin B_1 (AFB_1) on corn kernels. *Food Control* **2015**, *51*, 171–176. [CrossRef]

20. Pearson, T.C.; Wicklow, D.T.; Maghirang, E.B.; Xie, F.; Dowell, F.E. Detecting aflatoxin in single corn kernels by transmittance and reflectance spectroscopy. *Trans. ASAE* **2001**, *44*, 1247–1254. [CrossRef]

21. Dowell, F.E.; Pearson, T.C.; Maghirang, E.B.; Xie, F.; Wicklow, D.T. Reflectance and transmittance spectroscopy applied to detecting fumonisin in single corn kernels infected with *Fusarium verticillioides*. *Cereal Chem.* **2002**, *79*, 222–226. [CrossRef]

22. Fernández-Ibañez, V.; Soldado, A.; Martínez-Fernández, A.; De la Roza-Delgado, B. Application of near infrared spectroscopy for rapid detection of aflatoxin B_1 in maize and barley as analytical quality assessment. *Food Chem.* **2009**, *113*, 629–634. [CrossRef]

23. Singh, C.B.; Jayas, D.S.; Paliwal, J.; White, N.D.G. Fungal detection in wheat using near-infrared hyperspectral imaging. *Trans. ASABE* **2007**, *50*, 2171–2176. [CrossRef]

24. Zhang, H.; Paliwal, J.; Jayas, D.S.; White, N.D.G. Classification of fungal infected wheat kernels using near-infrared reflectance hyperspectral imaging and support vector machine. *Trans. ASABE* **2007**, *50*, 1779–1785. [CrossRef]

25. Bauriegel, E.; Giebel, A.; Geyer, M.; Schmidt, U.; Herppich, W.B. Early detection of Fusarium infection in wheat using hyper-spectral imaging. *Comput. Electron. Agric.* **2011**, *75*, 304–312. [CrossRef]

26. Williams, P.J.; Geladi, P.; Britz, T.J.; Manley, M. Investigation of fungal development in maize kernels using NIR hyperspectral imaging and multivariate data analysis. *J. Cereal Sci.* **2012**, *55*, 272–278. [CrossRef]

27. Yao, H.; Hruska, Z.; Kincaid, R.; Brown, R.L.; Bhatnagar, D.; Cleveland, T.E. Detecting maize inoculated with toxigenic and atoxigenic fungal strains with fluorescence hyperspectral imagery. *Biosyst. Eng.* **2013**, *115*, 125–135. [CrossRef]

28. Song, D.; Song, L.; Sun, Y.; Hu, P.; Tu, K.; Pan, L.; Yang, H.; Huang, M. Black heart detection in white radish by hyperspectral transmittance imaging combined with chemometric analysis and a successive projections algorithm. *Appl. Sci.* **2016**, *6*. [CrossRef]

29. Burns, D.A.; Ciurczak, E.W. *Handbook of Near-Infrared Analysis*, 3rd ed.; Taylor & Francis Group: Boca Raton, FL, USA, 2008.

30. Guiñón, J.L.; Ortega, E.; García-Antón, J.; Pérez-Herranz, V. Moving average and Savitzki-Golay smoothing filters using Mathcad. In Proceedings of the International Conference on Engineering Education-ICEE, Coimbra, Portugal, 3–7 September 2007.

31. Rinnan, Å.; van den Berg, F.; Engelsen, S.B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends Anal. Chem.* **2009**, *28*, 1201–1222. [CrossRef]

32. Chen, H.; Song, Q.; Tang, G.; Feng, Q.; Lin, L. The combined optimization of Savitzky-Golay smoothing and multiplicative scatter correction for FT-NIR PLS models. *ISRN Spectrosc.* **2013**, *2013*. [CrossRef]

33. Candolfi, A.; De Maesschalck, R.; Jouan-Rimbaud, D.; Hailey, P.A.; Massart, D.L. The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra. *J. Pharm. Biomed. Anal.* **1999**, *21*, 115–132. [CrossRef]

34. Lasaponara, R. On the use of principal component analysis (PCA) for evaluating interannual vegetation anomalies from SPOT/VEGETATION NDVI temporal series. *Ecol. Model.* **2006**, *194*, 429–434. [CrossRef]

35. Eriksson, L.; Andersson, P.L.; Johansson, E.; Tysklind, M. Megavariate analysis of environmental QSAR data. Part I—A basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). *Mol. Divers.* **2006**, *10*, 169–186. [CrossRef] [PubMed]

36. Himberg, J.; Mantyjarvi, J.; Korpipaa, P. Using PCA and ICA for exploratory data analysis in situation awareness. In Proceedings of the International Conference on Multisensor Fusion and Integration for Intelligent Systems, Baden-Baden, Germany, 20–22 August 2001; pp. 127–131.

37. Galvao, R.K.H.; Araujo, M.C.U.; Jose, G.E.; Pontes, M.J.C.; Silva, E.C.; Saldanha, T.C.B. A method for calibration and validation subset partitioning. *Talanta* **2005**, *67*, 736–740. [CrossRef] [PubMed]

38. Devos, O.; Ruckebusch, C.; Durand, A.; Duponchel, L.; Huvenne, J.P. Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation. *Chemom. Intell. Lab. Syst.* **2009**, *96*, 27–33. [CrossRef]

39. Chen, Q.; Zhao, J.; Fang, C.H.; Wang, D. Feasibility study on identification of green, black and Oolong teas using near-infrared reflectance spectroscopy based on support vector machine (SVM). *Spectrochim. Acta A* **2007**, *66*, 568–574. [CrossRef] [PubMed]

40. Guo, X.C.; Yang, J.H.; Wu, C.G.; Wang, C.Y.; Liang, Y.C. A novel LS-SVMs hyper-parameter selection based on particle swarm optimization. *Neurocomputing* **2008**, *71*, 3211–3215. [CrossRef]

41. Huang, C.L.; Dun, J.F. A distributed PSO–SVM hybrid system with feature selection and parameter optimization. *Appl. Soft Comput.* **2008**, *8*, 1381–1391. [CrossRef]

42. Brown, R.L.; Chen, Z.Y.; Cleveland, T.E.; Russin, J.S. Advances in the development of host resistance in corn to aflatoxin contamination by *Aspergillus flavus*. *Phytopathology* **1999**, *89*, 113–117. [CrossRef] [PubMed]

43. Pearson, T.C.; Wicklow, D.T. Detection of corn kernels infected by fungi. *Trans. ASABE* **2006**, *49*, 1235–1245. [CrossRef]

44. Manley, M.; McGoverin, C.M.; Engelbrecht, P.; Geladi, P. Influence of grain topography on near infrared hyperspectral images. *Talanta* **2012**, *89*, 223–230. [CrossRef] [PubMed]