JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

ARTICLE

pubs.acs.org/JAFC

Analysis of Pregerminated Barley Using Hyperspectral Image Analysis

Morten Arngren,[*,†,‡] Per Waaben Hansen,[*,‡] Birger Eriksen,[*,§] Jan Larsen,[*,†] and Rasmus Larsen[*,†]

[†]DTU Informatics, Technical University of Denmark, Richard Petersens Plads Bldg 321, DK-2800 Kgs. Lyngby, Denmark

[‡]FOSS Analytical A/S, Slangerupgade 69, DK-3400 Hillerød, Denmark

[§]Sejet Planteforædling I/S, Nørremarksvej 67, DK-8700 Horsens, Denmark

ABSTRACT: Pregermination is one of many serious degradations to barley when used for malting. A pregerminated barley kernel can under certain conditions not regerminate and is reduced to animal feed of lower quality. Identifying pregermination at an early stage is therefore essential in order to segregate the barley kernels into low or high quality. Current standard methods to quantify pregerminated barley include visual approaches, e.g. to identify the root sprout, or using an embryo staining method, which use a time-consuming procedure. We present an approach using a near-infrared (NIR) hyperspectral imaging system in a mathematical modeling framework to identify pregerminated barley at an early stage of approximately 12 h of pregermination. Our model only assigns pregermination as the cause for a single kernel's lack of germination and is unable to identify dormancy, kernel damage etc. The analysis is based on more than 750 Rosalina barley kernels being pregerminated at 8 different durations between 0 and 60 h based on the BRF method. Regerminating the kernels reveals a grouping of the pregerminated kernels into three categories: normal, delayed and limited germination. Our model employs a supervised classification framework based on a set of extracted features insensitive to the kernel orientation. An out-of-sample classification error of 32% ($CI_{95\%}$: 29−35%) is obtained for single kernels when grouped into the three categories, and an error of 3% ($CI_{95\%}$: 0−15%) is achieved on a bulk kernel level. The model provides class probabilities for each kernel, which can assist in achieving homogeneous germination profiles. This research can further be developed to establish an automated and faster procedure as an alternative to the standard procedures for pregerminated barley.

KEYWORDS: barley, pregermination, BRF method, near-infrared, NIR, hyperspectral camera, multinomial classification

INTRODUCTION

Barley is a cereal grain and is the cereal crop ranking number four in the world both in terms of quantity produced and in area of cultivation. Barley has many uses. It serves as a major animal fodder, as base malt for beer and certain distilled beverages, and as an ingredient in various food products. Only high quality barley is used for malting and beer production. Grain lots not fulfilling the quality requirements for malting will be downgraded to feed barley, which has lower economic value. The quality is defined by many different parameters, e.g. protein level, grading, dormancy, pregermination, physical damage in terms of inability to germinate, or even identification of various fungal infections. Identifying pregerminated barley in particular has so far been one of the most difficult quality parameters to handle both for the malting/brewing industry and for farmers/grain companies.

Pregermination (preharvest sprouting) is a major problem in seasons with wet harvest conditions. Germination will be initiated in the field and later stop again because of a change in weather conditions or drying of the grain after harvest and before storing. The barley may have little or no visible indication of the pregermination. The major defect of pregerminated barley is its predisposed inability to germinate again, e.g. for malting. When the germination reaches a certain level and is afterward stopped, the kernel can simply not continue the germination or will germinate with a reduced speed and vigor. A different limitation is kernels having dormancy as they cannot germinate due to a chemical lock preventing the germination process to start. This trait is hence sometimes considered as a positive trait as is hinders pregermination, but can require additional storing time to overcome the dormancy prior to malting.

Currently the germination ability of barley kernels can be evaluated using different approaches. The kernels can be subjected to a simple germination process run over a few days and with relatively few kernels by applying for instance the BRF method.[1] Specifically to determine pregermination, different staining methods can be used.[2,3] Visual inspection is also widely used, but it is subjective and lacks the sensitivity to detect sprouting early.[4] Common to these approaches is their high time consumption. Any method able to identify pregerminated barley within minutes or even classify the individual kernels into grades of quality will provide a tool for additional quality assessment to the industry. Basing such a system on computer vision provides the fast and nondestructive approach called for.

The research in pregerminated cereals has received attention in research communities to address the issue of identifying pregerminated barley kernels using automated visual systems. The simplest systems use cameras operating in the visual spectral range providing a high resolution. These image systems commonly rely on the spatial information and are often only able to detect germination by identifying a sprout on the kernels or to detect kernel deformations.[5] Takeuchi et al.[6] used such a computer vision system to detect morphological changes of barley kernels during germination based on 50 barley kernels.

The visual spectral range lies between 400 and 800 nanometers (nm) and has the limitation of only acting on the surface of

Received: May 31, 2011
Accepted: September 21, 2011
Revised: September 21, 2011
Published: September 21, 2011

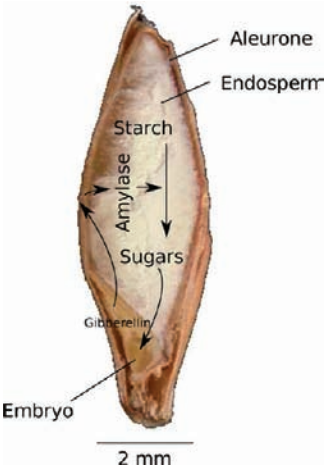

Figure 1. The internal biochemical mechanisms of a barley kernel during germination (courtesy of KAMPFFMEYER Food Innovation GmbH).

the kernels with insignificant penetration through the husk. Thus only identification of pregermination from external features is possible. As the initial process of germination is occurring inside at the embryo and endosperm, such an approach potentially leads to delayed detection. Spatial features used in visual systems can further be sensitive toward the orientation of the kernel.

A natural extension is to include near-infrared (NIR) wavelengths from about 800 to 2500 nm, where the light has the capability to penetrate the kernels through the husk and extract internal image information. NIR imaging thus has the potential to extract chemical changes before they cause an effect on the outside of the kernel. This approach can further lead to early detection of germination and can be superior to visual based systems. Xing et al.[5] combined the visual and NIR spectral ranges (400−1000 nm) to detect sprout damage to Canada Western Red Spring wheat. They conducted classification on approximately 200 wheat kernels germinated for 0, 24 and 48 h by exploiting various spatial changes of the kernels during germination.

Many applications using the visual and NIR spectral range are based on fewer than 20 different wavelengths. This spectral resolution may not be sufficient in applications, where similar constituents need to be separated. Hyperspectral imaging provides a high spectral resolution with more than 100 bands, usually at the expense of a lower spatial resolution. This allows for models to utilize the spectral information, possibly in combination with the image data. Xing et al.[7] exploited this combination to classify α-amylase activity in individual Canadian western wheat kernels into two classes using hyperspectral NIR images. The predictions were based on the partial least-squares (PLS) model using mean spectra from the image segmented embryo leading to classification rates above 80%.

Munck and Møller[8] had developed a model based on NIR spectroscopy with separately acquired image information to develop a model able to predict the 24 h germination of barley kernels. Engelbrecht et al.[9] illustrated the potential of using hyperspectral imaging from 1000 to 2500 nm to classify barley into normal or germinated kernels based on 150 grains germinated for 6, 9, 12, 18 and 24 h. Their work further included the destructive tetrazolium test to obtain single kernel reference values on their viability. Manley et al.[10] extended the work to

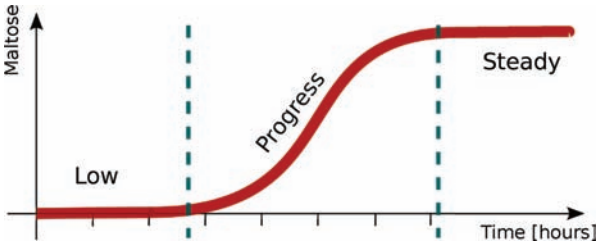

Figure 2. Conceptual illustration of the progression of germinating barley showing the starch breakdown. The two thresholds indicate the start and stop of the major chemical changes inside the kernel during germination.

employ a PLS discriminant analysis model to classify single kernels into normal or pregerminated and achieve a classification error of 2.87%.

Analyzing barley using NIR imaging by building a mathematical model to provide the degree of pregermination of single kernels with true pregermination times as references has not been attempted previously. We therefore investigate if a mathematical model can be developed for the detection of pregerminated barley and for describing the degree of pregermination of a single kernel using hyperspectral NIR imaging with true pregermination time as reference.

In the next section we give an introduction to barley germination in general, including the pregermination experiment followed by a description of the camera system, the preprocessing pipeline and the theoretical background for our analysis and model choice. The classification results are presented and discussed in the Results. A summary with discussions is presented in the last section. Supplementary details concerning the image acquisition and the models used are described in a supplementary technical report.[11]

MATERIALS AND METHODS

Barley Germination. A barley kernel (*Hordeum vulgare*) is a member of the grass family and has internal structure as illustrated in Figure 1. Germination of barley is the process of the grain starting to sprout and becoming a plant. Biologically the process is highly complex with many stages, and in this context only the general steps will be considered. It partly involves the conversion of starch into sugars (maltose and glucose) used for the growth. The germination process is initiated by the absorption of water into the kernel. Provided the environmental conditions are suitable, e.g. temperature, this starts a whole series of biological steps including the release of the enzyme α-amylase from the aleurone layer into the endosperm, as shown in Figure 1. The enzyme breaks the starch molecules down into oliosaccharides, subsequently to the disaccharide maltose, and finally to glucose, which is transported to the embryo for radicle and epicotyl growth. This activity can be seen as a large scale physical change inside the kernel. A conceptual illustration of the progression of the germination in barley is depicted in Figure 2. It starts with a short period with little activity, where the α-amylase release is commencing, then rises as the amylase activity is maximum and finally settles as all the starch is completely broken down.

Pregermination Experiment. An experiment was set up in collaboration with Sejet Planteforaedling, Denmark, to produce barley kernels of the Rosalina variety with eight different durations of pregermination from 0 to 60 h. The experiment was conducted using the following procedure.

Table 1. Number of Kernels Used in the Different Steps in the Analysis and the Concentration of Water in the Kernels with an Accuracy of 0.25%age Points[a]

	0 h[b]	12 h[b]	18 h[b]	24 h[b]	30 h[b]	36 h[b]	48 h[b]	60 h[b]	total
pregerminated	2000	500	500	2000	2000	2000	500	500	10000
regerminated	300	300	300	1000	300	300	300	300	3100
image acquired	150	97	50	150	150	150	47	46	840
preprocessed	135	87	45	135	135	135	42	41	755
water	10.4%	10.2%	9.9%	10.2%	10.2%	10.1%	10.0%	9.8%	

[a] Since not all germination times had the same number of kernels, only images of a subset of the kernels were acquired. [b] Germination time.

1. Two filter papers (upper and lower side) were soaked in water for a few hours and afterward left to drip off. Prior to use, the filter paper must still be wet, but not soaking or dripping.
2. 100 single kernels were placed on the lower filter paper.
3. The upper filter paper was placed on top and the enclosed kernels were rolled up to a small cylinder locked with a rubber band and placed in a dark and warm environment (e.g., a cardboard box at room temperature).
4. When the germination times were reached, the individual kernels were immediately put in a net bag and placed in front of a blow dryer for a minimum of three hours at 5 to 10 °C.

500 kernels were pregerminated for each of the times [12, 18, 48, 60 h], and 2000 kernels were pregerminated for each of the times [0, 24, 30, 36 h] in order to conduct a reliable statistical analysis in the time range, where the kernels would start and stop the major chemical change (cf. Figure 2). When the moisture level inside the kernels drops below approximately 35% during the drying process, the germination process is considered halted. The kernels were further dried to a water content of approximately 7−27%, which is considered as a stable level for long-term storage.

After the kernels had pregerminated and had been dried by blow drying, they were stored in a refrigerator for approximately 6 months. Over 3000 of the kernels were then subjected to a regermination experiment (described below) to measure the degrading effect of the pregermination. Including wastage this left 6200 kernels for an additional 4 months of storage before the hyperspectral images were taken. Since not all germination times had the same number of kernels, only images of a subset of the kernels were acquired, i.e. 840 kernels (cf. Table 1). This long-term storage slowly dried the kernels further and led to equal water contents of approximately 10% between the pregermination times using FOSS NIR XDS instrument analysis (cf. Table 1).

Prior to the hyperspectral image acquisition, the epicotyl and radicle were rubbed off to fit the kernels on the sample plate. This has no impact on our analysis as we attempt to capture the pregermination inside the kernels using penetrating light (NIR) instead of relying on visible physical changes. During the image acquisition, the germination of every kernel was verified by visual inspection. This is, however, associated with some uncertainty as germination may not be visible for the short germination times. The germination time references designate the period the kernels were inside the germination environment, i.e. from when the filter paper is locked with rubber bands to the start of the dry blowing process. The periods obviously become a bit longer, as the germination will only halt when the water content falls below approximately 35%. This additional germination time is considered equal for all periods and is hence not included in the germination time references.

Germinative Energy of Barley. The degrading effect of pregermination can be measured by evaluating the germinative ability of the stable pregerminated barley kernels. This is done by subjecting the pregerminated and dried kernels to the standard BRF method,[1] where the kernels are regerminated. It consists of placing 100 kernels in a germinative environment (enclosed Petri dish with filter paper and water) and counting the sprouted kernels every 24 h for three days. The germinate energy is then the fraction of germinated kernels out of the 100 kernels. In our case the germinated kernels are counted for five days, using the same conditions as the BRF method.

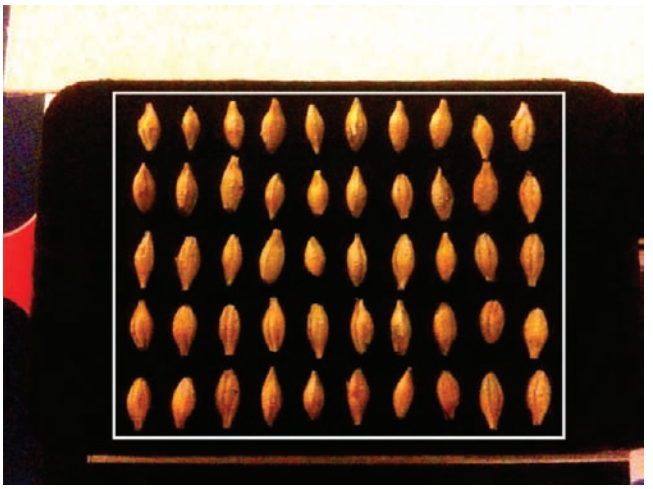

Figure 3. Barley kernel sample plate painted with a special NIR black color. A small strip of the NCS-0300 paper is placed on the top as the white reference acquired for each image. Afterward the kernels are cropped, as marked by the white rectangle.

Image Acquisition and Preprocessing. *Camera System and Equipment.* A hyperspectral line-scan NIR camera system from Headwall Photonics Inc. (Fitchburg, MA, USA) sensitive in the range 900 to 1700 nm was used to acquire the image of the barley kernels. A dedicated NIR light source illuminates the sample uniformly along the scan line, and an advanced optical system disperses the NIR light onto the InGaAs sensor inside the camera for acquisition. A sledge from MICOS GmbH (Germany) moves the sample past the view slot of the camera allowing it to acquire a hyperspectral image, line by line. The camera has a resolution of 320 spatial pixels and 166 spectral pixels with equally spaced wavelength intervals of 4.8 nm. Further details of the performance of the camera system and the standard operating procedure can be found in the technical report, cf. ref 12.

Acquisition Procedure. The procedure for acquiring images can be split into an initialization part to set up the camera system and a subsequent acquisition part. The initialization procedure consists of a series of steps with the purpose to acquire maximum quality images of the barley kernels. Initially the camera system and light source are tuned to maximize dynamic range while limiting specular reflections and saturated pixels. Images of the wavelength standard WSR1920a[13] from Avian Technologies LLC (Sunapee, NH, USA) and a set of vertical lines were acquired for wavelength scale alignment and spatial calibration respectively. The entire process of acquiring images of all the barley kernels once was conducted in one run in a single day.

In order to obtain usable images a maximum of 50 barley kernels were fixated at a time on a sample plate, as shown in Figure 3. The sample plate (courtesy of Skandinavisk Bryggeri Laboratorium, Denmark) is painted with a NIR dark black color (Mankiewicz Nextel-Velvet Coating 811-21 9218 Schwartz) to avoid reflections during acquisitions and for easier background removal during data processing. The acquired images include a white reference [NCS-0300 (Natural Color System, Sweden, Web: http://ncscolour.com) reflecting paper] as the top part in every image to ensure the capture of any potential drift of the light source. An equivalent dark current image is taken at the end of the acquisitions for

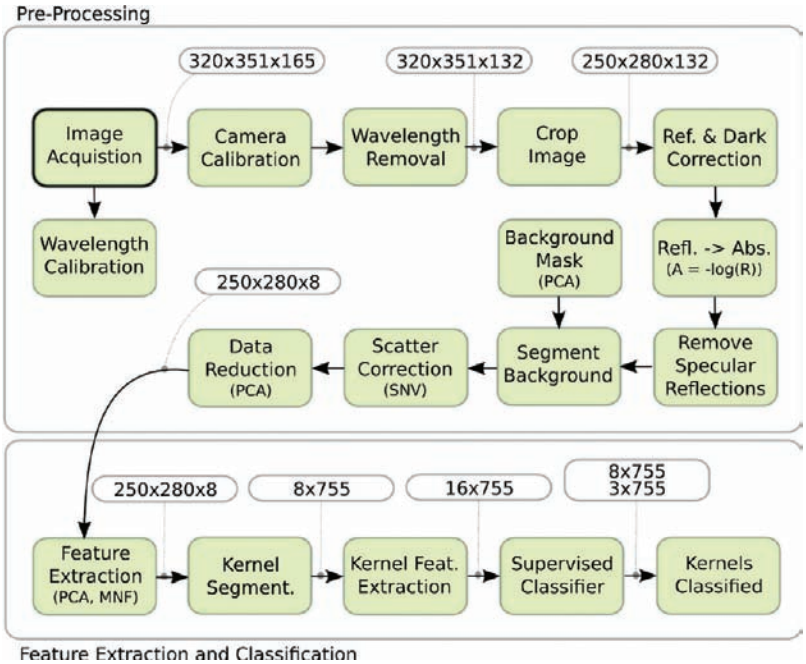

Figure 4. Flowchart of the preprocessing, feature extraction, and classification steps starting from the boldface *Image Acquisition* box. The pinned boxes show the size of the image data at the different steps.

each germination time by obstructing the incoming light to the camera with a lens cap.

Preprocessing Pipeline. The preprocessing pipeline, as illustrated in Figure 4, is made from a series of steps to prepare the data structure for subsequent analysis. Initially the wavelength scale is recalculated using the wavelength standard WSR1920a[13] to achieve optimal calibration. The camera system suffers from a small misalignment in the optical system, which causes a minor spatial distortion over both the spectral and spatial range. This is corrected using an image of a set of vertical lines and an image of the wavelength standard WSR1920a for calibration. The wavelength range is reduced to 132 bands between 1002 and 1626 nm due to the poor sensitivity of the InGaAs sensor at the spectral extremities. The camera further suffers from a single vertical line of poor sensitivity across the first column of barley kernels, and the affected pixels will represent outliers. These kernels are eliminated from the data set in all images, leaving only 45 kernels per image and a reduced data set of 755 kernels, cf. Table 1. Each acquired image is then represented in a 250 pixels × 280 pixels × 132 bands 3-way tensor after cropping the kernels out (white rectangle in Figure 3). Every image is corrected with the associated white reference and dark current image by $R = (I_{raw} - I_{dark})/(I_{white} - I_{dark})$. The spectral mixing is assumed dominated by subsurface penetration and thus nonlinear multiplicative mixing. To enforce linear mixing the spectra are transformed from reflectance data to absorbance data by $A = -\log_{10}(R)$. Pixels suffering from specular reflections do not hold any spectral information relevant for the analysis. They are identified with a reflection larger than one $R > 1$ or a negative absorbance $A < 0$ and are removed. The background can easily be removed due to the high contrast between the kernels and the NIR absorbing black color by a simple approach. This contrast will exhibit the highest variance and can be captured by a principal component analysis (PCA) decomposition. This is conducted by unfolding each tensor image data into a matrix, where each column represents a pixel, and subjecting the mean centered matrix to PCA. The foreground and background contrast is represented in the first score image and is used to extract a mask from a simple scalar threshold. The acquired images are further corrected for scatter effects as a preprocessing step on the spectral information. Many different scatter correction techniques exist, and we will apply the *standard normal variate* (SNV) correction.[14] Each of the acquired spectra consists of 132 bands, and the relevant information for our analysis is expected to be represented by fewer components. The spectral information is therefore compressed to D dimensions, which is set manually, to suppress the noise allowing for a reduction in the computational load and curse of dimensionality, but at the risk of removing information important for the final analysis. In order to represent all germination times a concatenated data subset is formed from one SNV corrected image from each germination time. This tensor subset is unfolded to a matrix, where each column holds a spectrum. The mean centered data is subjected to PCA, and the entire data set is projected onto the first D loadings.

Classification Models. The classification of the individual kernels is based on both spectral and spatial features from the image data, and these features are extracted prior to the classifier (cf. Figure 4).

Feature Extraction. Prior to the actual classification, the image data is processed to find features, which exhibit the chemical changes inside the kernels related to the germination process. The feature extraction is applied to a subset of the compressed images represented by one image from each germination time to capture these chemical changes. To this end several feature extraction methods exist, and we will focus on two simple approaches: *PCA*[15] and *minimum noise fraction (MNF)*.[16] PCA finds a linear combination of the spectra, which maximizes the variance in the data and hence does not consider any spatial information. These features have already been extracted in the data compression step and are reused. MNF also finds a linear combination of the spectra by maximizing the signal-to-noise ratio. The signal covariance matrix is based on the spectral information, and the noise covariance matrix is estimated from a neighbor pixel dependency. It thus exploits both spectral and spatial information to estimate components to explain more variance in fewer components than PCA does. The extracted features from any of the decompositions are finally represented as a set of score images.

Single Kernel Segmentation and Feature Extraction. The extracted features reveal the chemical changes inside the kernels during germination and need to be quantified prior to the classification. Initially the kernels are segmented out easily since the kernels have the same position

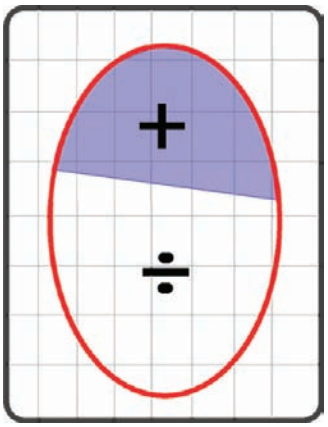

Figure 5. The single kernel feature consists of two scalar values for each score image calculated as the sum of the positive and negative intensities respectively divided by the total number of pixels.

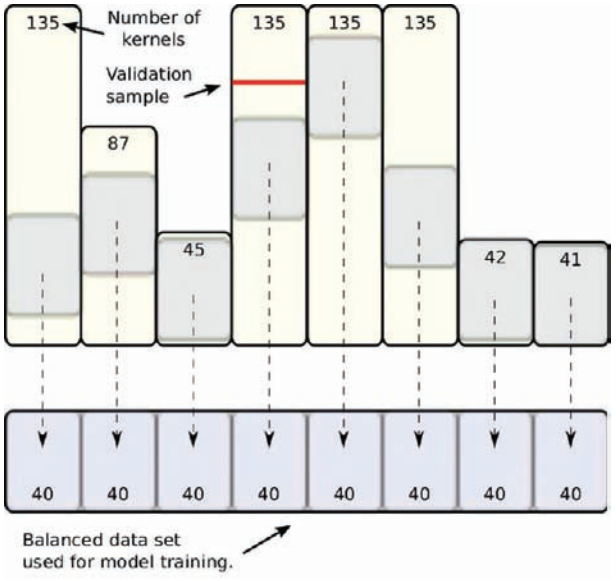

Figure 6. Leave-one-out cross validation framework with subsampling to extract training/test sets with equal number of samples from each germination time.

on the sampling plate for all images. In order to quantify the germination progression, a second level of features are extracted from each of the kernels. This feature is based on the spatial information in the extracted score images, as shown in Figure 5, and is constructed to maximize the contrast between the starch break down progression and the background. The feature can be described as the sum of the pixel intensities for each score image for each barley kernel divided by the total number of pixels in the kernel for both positive and negative intensities separately. The advantage of this feature is the insensitivity to the kernel orientation. If we denote the extracted score images $\mathcal{X}^{\text{score}} \in \mathbb{R}^{X\times Y\times D}$, where X and Y are the image dimensions and D is the number of loadings, then the single kernel feature $\mathbf{f}$ can be expressed as

$$\mathbf{f}_{k,n} = \frac{1}{L}\left[\sum_{x_+} x_{ijk}^{\text{score}} \sum_{x_-} x_{ijk}^{\text{score}}\right] \quad (1)$$

where k represents the kth band, n represents the nth kernel, i,j denotes the pixel position and L is the total number of pixels in the kernel. $\mathbf{f}_{k,n}$ is

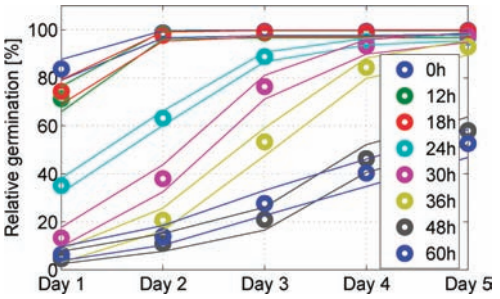

Figure 7. Progression of regerminated barley for each day showing the accumulated percentage of germinated kernels for each pregermination duration (dots) with 95% confidence intervals (lines). The progression reveals three clusters of pregermination: normal, delayed and limited.

calculated for each score image k for each kernel n as a 2-dimensional vector, which leads to a complete feature vector $\mathbf{F}_n = [\,\mathbf{f}_{1,n}, \mathbf{f}_{2,n}, ..., \mathbf{f}_{D,n}\,]$ of length $2D$ for each kernel n.

Classification. We have evaluated different classification models based on different approaches, which includes ordinal classifiers and nonlinear neural networks (cf. ref 11 for supplementary material). The model with the best performance is the maximum likelihood multinomial regression classifier implemented in the L1General toolbox (available for download at http://pages.cs.wisc.edu/gfung/GeneralL1),[17,18] which provides a probability for each barley kernel n for each class c (cf. Figure 4). The classifier can be expressed by the softmax function

$$p(c|\mathbf{F}_n) = \frac{\exp(\mathbf{F}_n\mathbf{w}_c)}{\sum_{\hat{c}=1}^{C}\exp(\mathbf{F}_n\mathbf{w}_{\hat{c}})} \quad (2)$$

where the model weights are denoted by $\mathbf{w}_c$ for each class c, and C denotes the total number of classes. The entire data set is applied in a leave-one-out cross validation framework, where the training and test sets are extracted from the entire data set as illustrated in Figure 6 to ensure an equal number samples from each germination time. A different training and test set combination can be extracted repeatedly to achieve an average class probability for each validation sample. In commercial applications bulk results are often desired. Single kernel class probabilities are hence afterward combined as the average class probabilities of adjoining kernels to achieve a more robust probability and provide bulk classification results. A confusion matrix for all C classes is calculated using the class probabilities to evaluate the classification performance. The C classes can be aggregated to a smaller confusion matrix with fewer classes, $\hat{C}$, or the model can be retrained with $\hat{C}$ classes to obtain a more dedicated model with potentially lower classification error. In our case the former approach is used to model the $\hat{C}$ classes.

RESULTS

This section describes the model and classification results. Supplementary details can be found in the associated technical report.[11]

Germinative Energy of Barley. The degrading effect of pregermination is evaluated by subjecting stable pregerminated barley kernels to the standard BRF method,[1] as described previously. The BRF method was repeated several times for each pregermination duration for improved accuracy (cf. Table 1). The resulting progression of the regermination is shown in Figure 7. The regermination results suggest the pregerminated barley can be divided into three groups:

1. Normal regermination. A short pregermination has negligible effect as all kernels regerminate as normal kernels. The period includes 0 to 18 h of pregermination.

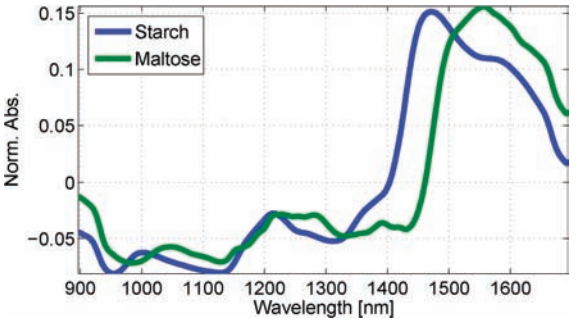

Figure 8. NIR spectra of pure constituents of starch and maltose in powder form using our hyperspectral NIR camera system. The two spectra exhibit large differences in the spectral range from 1400 to 1550 nm.

2. Delayed regermination. The kernel will regerminate within five days, but at a slower rate. This period is between 24 to 36 h of pregermination.
3. Limited regermination. The kernels are exhausted from pregermination and not all will regerminate within the five days. This period is greater than 36 h of pregermination.

The division further indicates how the germination starts a major chemical change between 18 and 24 h. Applying the BRF method to pregerminated kernels with visual radicle and epicotyl may fail as the method relies on visual inspection. Long pregermination durations will lead to a visible radicle and epicotyl, and hence identifying if the individual kernel has in fact germinated again becomes difficult. However, this does not change the three pregermination categories identified.

Preprocessing. The acquired image data is preprocessed as described in the Preprocessing Pipeline section.

Background Segmentation and Scatter Correction. The barley kernels are segmented out by using the simple procedure described earlier. Each hyperspectral image is unfolded into a 132 $\times$ N matrix, where N = 250 $\times$ 280 pixels $\approx$ 70000 (excluding specular reflecting pixels). The background mask is extracted from the first loading from the PCA decomposition. After the background removal each kernel is on average represented by approximately 400 pixels. Each spectrum is afterward scatter corrected by applying the SNV transform.[14] A detailed comparison with other scatter correction methods is described in the associated technical report,[11] which shows how the SNV method is superior to the others in this application.

Data Compression. The data is compressed as described earlier. The mean centered data subset is represented in a matrix of size 132 $\times$ $N_{samples}$, where $N_{samples}$ = 8 germination durations $\times$45 kernels per image $\times$ $\sim$400 pixels per kernel $\approx$ 140000 spectra, and subjected to PCA. The number of PCs required to capture the most important information is evaluated from a set of estimated projection loadings. To ensure that the relevant information is retained in the data, while achieving a suitable compression, the entire data is projected onto the first D = 8 PC's. These components explain 99% of the variance.

Data Visualization. For evaluation of NIR spectra, hyperspectral images of the main constituents during germination, starch and maltose in their pure powder form, were acquired using our camera system, and their mean spectra are illustrated in Figure 8. The starch and maltose spectra reveal similar variation in the lower part of the spectrum, but exhibit large differences at the longer wavelengths. We therefore expect to observe spectral activity during germination in the area between 1400 and 1550 nm. The hyperspectral image data of the barley kernels can be visualized from the score images and the associated spectral loadings from the PCA decomposition conducted in the data compression step. The resulting score images can be used to analyze the internal structure of the barley kernels in general and during germination. Figure 9 illustrates the first five score images and spectral loadings of the second and third PC. The second spectral loading is comparable to the spectra of starch and maltose. Activity at the embryo is clearly captured by the third and fifth score images. The third spectral loading reveals a clear peak around 1430 nm indicating spectral activity in the same area as the spectral difference between starch and maltose. The husk of the barley kernel acts as an attenuating filter and is observed as an intensity difference. However, this is not visible due to the normalizing effect of the SNV scatter correction. The fifth score images for the germination times [0, 12, 24, 36, 48 h] are shown in Figure 10. This PC also captures the germination progression identified as the small area at the embryo increasing over time. The corresponding spectrum includes additional spectral features to increase the contrast of the germination progression compared to PC loading 3. As the third and fifth components designate the starch breakdown into maltose, their associated spectral loadings represent the largest contrast between starch and maltose and not necessarily a pure spectral signature.

A visual subjective evaluation based on several similar images indicates a threshold of approximately 24 h, where the kernels would start the major chemical change. Similarly the same images reveal a very small difference between 48 and 60 h indicating the steady state and hence the second threshold between 36 and 48 h. These properties of germination coincide with the results from the germinative energy analysis using the BRF method.

Feature Extraction. The PCA features have already been calculated in the data compression step, and the 8 score images are hence reused. For the MNF decompositions the compressed data subset is represented in a matrix of size 8 $\times$ $N_{samples}$, where $N_{samples}$ = 8 germination durations $\times$ 45 kernels per image $\times$ $\sim$400 pixels per kernel $\approx$ 140,000 spectra. Features from the entire data set are then extracted using the estimated loading vectors for each method. The MNF method extracts similar components compared to PCA as shown in Figure 11, where the germination progression is clearly captured in component 1. The germination progression is found in a single component as opposed to PCA leading to a stronger contrast between the germinating embryo and the rest of the kernel. The associated spectrum designates the correlations between the MNF feature loading 1 and the original variables. It is comparable to the fifth PC loading in Figure 10 as it includes similar spectral features to increase the contrast of the germination progression. In general the MNF is considered to perform best in terms of extracting spatial features for the classification. The extracted features used in the following are therefore 8 score images from the MNF decomposition.

Single Kernel Segmentation and Feature Extraction. The kernels are segmented out and the single kernel feature to quantify the germination progression $\mathbf{F}_n$ is calculated for each kernel n leading to a 2 $\times$ 8 = 16-dimensional feature vector per barley kernel. For all barley kernels, the entire feature extracted data set $\mathbf{F}$ becomes a 16 $\times$ 755 matrix.

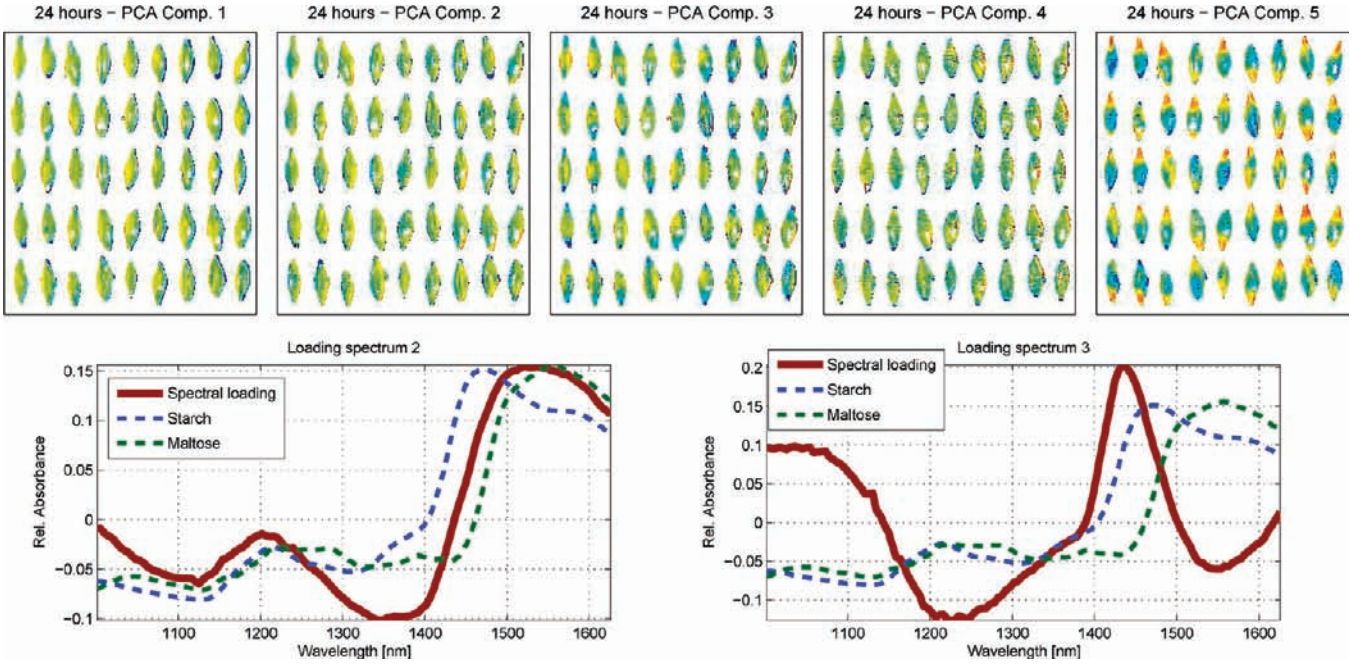

Figure 9. PC score images with the corresponding spectral loadings for PC2 and PC3. The second PC loading is comparable to the scaled pure spectra for starch and maltose acquired using the same camera system for comparison (dashed). Both the third and fifth score images have captured activity at the embryo indicating the germination process. The third spectral loading exhibits a distinct peak around 1430 nm and indicates activity in the same area as the spectral signatures of starch and maltose.

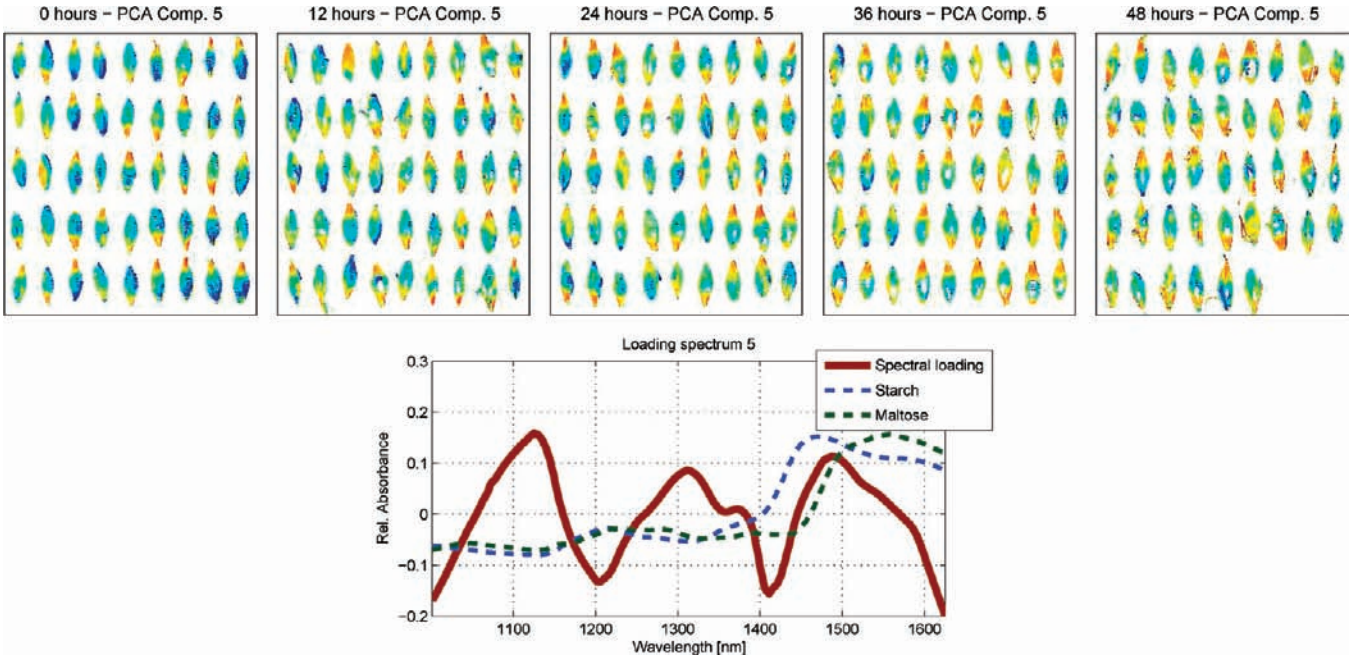

Figure 10. PCA score image 5 for the germination durations [0, 12, 24, 36, 48 h] (left to right) and the associated spectral loading. The score images reveal a clear progression in the germination process, and the spectral loading includes additional spectral features to increase the contrast of the germination progression compared to PC loading 3.

Classifiers. The classification on a single kernel level will be conducted into the eight classes designating the germination times [0, 12, 18, 24, 30, 36, 48, 60 h] using the MNF feature set.

The leave-one-out approach is applied by extracting a training and test set to ensure equal number of samples from each germination time, i.e. a balanced data set (cf. Figure 6). The germination time with fewest data is 60 h and has only 41 samples (cf. Table 1). Excluding the validation sample the balanced training/test set is hence set to include 40 samples from each germination time, i.e. a total of 320 samples.

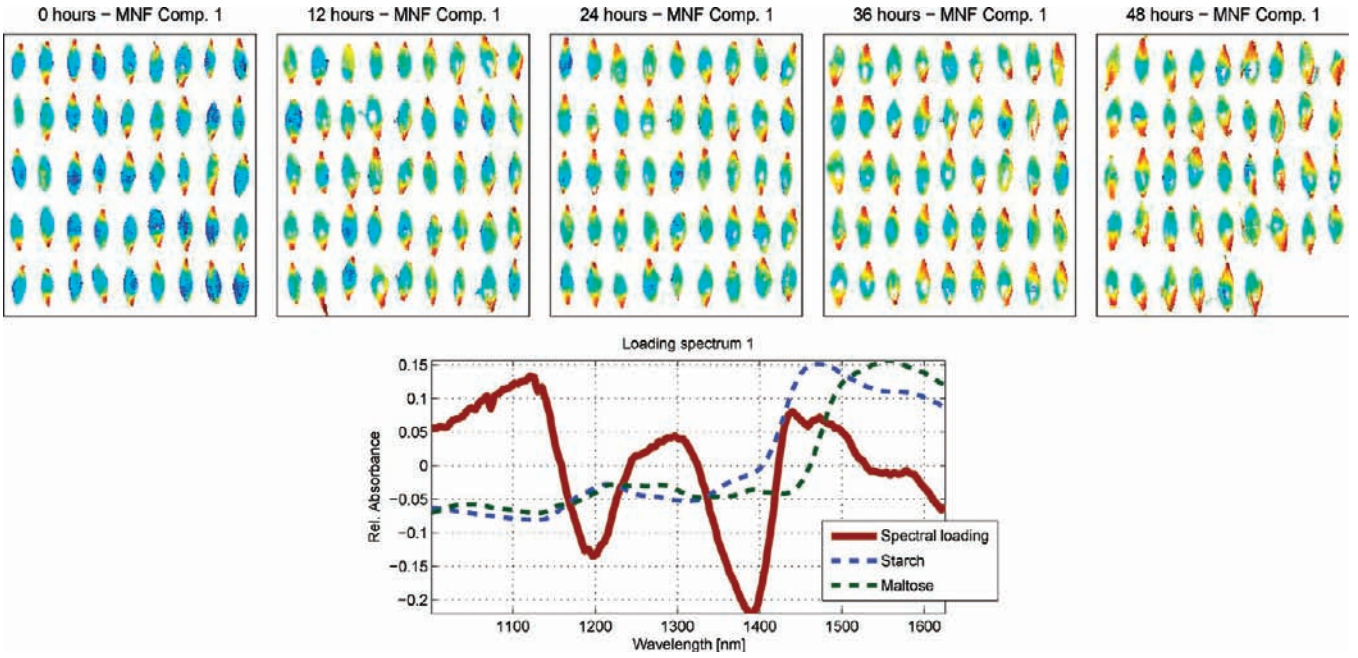

Figure 11. The MNF score image 1 on top represents the germination durations [0, 12, 24, 36, 48 h] (left to right) and reveals a stronger intensity of the germinating embryo than PCA does. The spectrum represents the correlations between the MNF feature vector 1 and the original variables. It exhibits spectral activity in the area, where the spectra of starch and maltose differ, and is comparable with PC loading 5 shown in Figure 9.

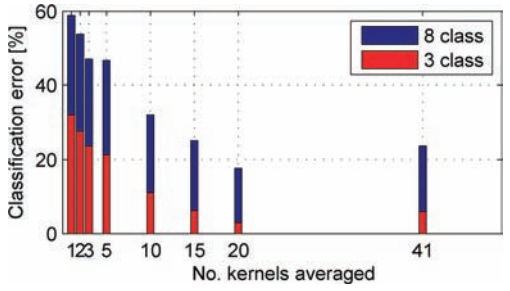

Figure 12. Classification errors for different numbers of averaged kernels. The error clearly decreases when averaging more kernels. Averaging kernels leaves fewer samples to analyze and hence higher sensitivity to our results. When averaging 41 kernels, the error is based on only 17 samples.

A random training/test set is extracted 3000 times for each validation sample, a corresponding 8-class classifier model is inferred for each training/test set, and the average class probability for each kernel is used. The classification distribution is calculated in an 8 × 8 confusion matrix to evaluate the classification performance and a corresponding 3 × 3 confusion matrix for the three classes. The validation classification error for all single kernels in all eight classes become 59% ($CI_{95\%}$: 55−62%) and 32% ($CI_{95\%}$: 29−35%) for the three aggregated classes. This can be compared to the random guess error rate of 7/8 = 87.5% and 2/3 = 66.7% respectively.

The class probabilities for adjoining kernels in each class are averaged to estimate bulk classification results for a small range of combinations as shown in Figure 12. The classification error clearly decreases and is a minimum by averaging 20 kernels with an error of 18% ($CI_{95\%}$: 7−35%) and 3% ($CI_{95\%}$: 0−15%) for eight and three classes respectively. Since we only have 755 kernels, averaging them leads to even fewer samples and less accurate results. This is evident for averaging 41 kernels, which only have 17 samples and a higher classification error.

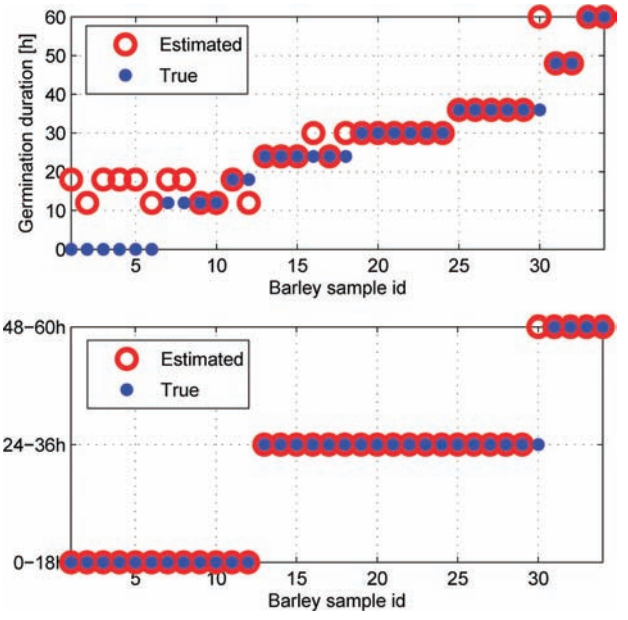

Figure 13. Error distribution for averaging 20 kernels for both eight and three classes (top and bottom respectively). The few misclassifications for the eight classes are concentrated around 0−12 h, but stay within the three aggregated classes. The resulting classification error for the three aggregated classes of 3% is due to one misclassification.

The classification distributions for eight and three classes for 20 averaged kernels are illustrated in Figure 13. The misclassifications for the eight classes are concentrated around 0 to 12 h, and are due to the pregermination process having not yet produced macroscopical changes. From this point of view the

pregermination process is with the same stage between 0 and 18 h of pregermination, and causes these misclassifications. The three class validation error of 3% corresponds to one misclassification out of 34 samples and is thus sensitive to the number of samples.

DISCUSSION

In summary our modeling framework based on hyperspectral NIR imaging technology has proven feasible for describing the degree of pregermination of single barley kernels. The approach of using NIR imaging technology has also proven superior to visually based systems in terms of visualizing the chemical changes inside the kernels during the germination process of barley kernels. The practical germination experiment with eight pregermination levels identified three groups to categorize pregerminated barley: normal, delayed and limited. We presented a simple classifier model to classify the germination durations for single barley kernels. By averaging single kernels to obtain bulk level results and aggregating to three classes we achieved a classification error of 3% ($CI_{95\%}$: 0−15%). Our results are only based on 755 kernels and can be revalidated using a larger data set.

Part of the framework development included evaluation of different approaches, i.e. compression rate and feature extraction method. These choices were made manually and should instead be taken by optimizing the complete classifier framework for a more comprehensive approach. In real-world applications the distribution of single kernels into the three classes is of interest. This gives a more nuanced profile of the pregermination of the kernels than their bulk average. In such applications the classification is performed on a single kernel level and the error of 32% is thus applicable. Our results suffers from having relatively few kernels in the analysis and can be made more accurate and reliable in a future experiment with a larger sample. Our model only assigns pregermination as the cause for a single kernel's lack of germination and is unable to identify dormancy, kernel damage etc. The focus of future work can therefore be to attemp to identify these other properties based on dedicated barley kernel experiments. This can be done in conjuction with our model to build an even more comprehensive framework. The measured water levels of the stored kernels were equal and can be considered a confounder. The impact in the prediction performance should be investigated empirically with other water concentrations. A more robust and difficult approach is to have different water levels for all single kernels to ensure that the model becomes independent of the water levels.

Using hyperspectral images provides the opportunity to achieve a spectrum with an abundance of bands for each pixel. This approach is suitable for research as in this case, but may not be optimal for commercial interests. A natural extension to this work would include the identification of the most important wavelengths to simplify the image technology. If similar results can be achieved with only a few wavelengths, then cost-effective multispectral image equipment can be used instead. Our framework is applicable to other germinating grain types and can possibly be used for other applications such as kernel fungal infections or perhaps detection of dormancy. It is currently based on a single variety of barley and can also be validated against other varieties for possible reinference of model parameters.

Compared to the standard BRF procedure to analyze barley kernels for pregermination taking several days, our proposed automated method is faster and does not include any subjective evaluation. A portion of kernels can be classified within a few minutes. With the system used in this research, image acquistion takes approximately 2 min per sample plate of 50 kernels and the export and classification of data is conducted within tens of seconds. In total the kernels can be analyzed within a few minutes.

The entire analysis is implemented and conducted in MATLAB with an associated toolbox available for download.[19]

AUTHOR INFORMATION

Corresponding Author

*E-mail: ma@imm.dtu.dk, moa@foss.dk pwh@foss.dk; bee@sejet.com; jl@imm.dtu.dk; rl@imm.dtu.dk.

ACKNOWLEDGMENT

The authors thank Sejet Planteforædling I/S for their assistance in the barley pregermination experiment, Skandinavisk Bryggeri Laboratorium, Denmark, for the sample plates and finally Bo Büchmann (FOSS Analytical A/S) for setting the entire cooperation up and the early review.

REFERENCES

(1) Germinative Energy of Barley: BRF Method. *Analytica-EBC*; 1997, *Section 3*, Method 3.6.2.

(2) Germinative Capacity of Barley: Rapid Staining Method. *Analytica-EBC*; 2005, *Section 3*, Method 3.5.1.

(3) Pre-Germinated Grains in Barley: Flourescein Dibutyrate Method. *Analytica-EBC*; 2006, *Section 3*, Method 3.8.1.

(4) Bason, M. L.; Ronalds, J. A.; Wrigley, C. W.; Hubbard, L. J. Testing for sprout damage in malting barley using the rapid visco-analyser. *Cereal Chem.* **1993**, *70* (3), 269–272.

(5) Xing, J.; Symons, S.; Shahin, M.; Hatcher, D. Detection of sprout damage in Canada Western Red Spring wheat with multiple wavebands using visible/near-infrared hyperspectral imaging. *Biosyst. Eng.* **2010**, *106* (2), 188–194.

(6) Takeuchi, R.; Kojima, H.; Toyoda, K.; Omoto, H.; Morimoto, S. Morphological measurement of barley seed by computer vision system measurement of changes in seed shape during germination. *Sci. Rep. Fac. Agric. Kobe Univ.* **1992**, *20* (1), 115–121.

(7) Xing, J.; Hung, P. V.; Symons, S.; Shahin, M.; Hatcher, D. Using a Short Wavelength Infrared (SWIR) hyperspectral imaging system to predict alpha amylase activity in individual Canadian western wheat kernels. *Sens. Instrum. Food Qual. Saf.* **2009**, 3 (4), 211–218.

(8) Munck, L.; Moller, B. A new germinative classification model of barley for prediction of malt quality amplified by a near infrared transmission spectroscopy calibration for vigour "on line" both implemented by multivariate data analysis. *J. Inst. Brew.* **2004**, *110* (1), 3–17.

(9) Engelbrecht, P.; Manley, M.; Williams, P. J.; Toit, G. D.; Geladi, P. Pre-germination detected in whole cereal grains using near infrared hyperspectral imaging. *Proc. CST SA—ICC Int. Grains Symp., Qual. Saf. Grain Crops Foods* **2010**, 123–127.

(10) Manley, M.; Engelbrecht, P.; Geladi, P. NIR hyperspectral imaging at Stellenbosch University: detection of pre-germination in whole sorghum grains. *NIR News* **2010**, *21* (7), 11–13.

(11) Arngren, M.; Larsen, J.; Hansen, P. W.; Eriksen, B.; Larsen, R. *Analysis of Pre-Germinated Barley using Hyperspectral Image Analysis, Technical Report*; http://www.imm.dtu.dk/pubdb/p.php?5989; Technical Report, 2010.

(12) Arngren, M. *Hyperspectral NIR Camera System, v1.1*; Technical Report, 2011

(13) Standard Reference Material 1920a (Near Infrared Reflectance Wavelength Standard from 740nm to 2000nm). National Institute of Standards and Technology (NIST).

(14) Barnes, R.; Dhanoa, M.; Lister, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* **1989**, *43* (5), 772–777.

(15) Hotelling, H. Relations between two sets of variables. *Biometrika* **1936**, *28* (3/4), 321–377.

(16) Green, A. A.; Berman, M.; Switzer, P.; Craig, M. D. A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal. *IEEE Trans. Geosci. Remote Sens.* **1988**, *26*, 65–74.

(17) Schmidt, M.; Fung, G.; Rosales, R. Fast optimization methods for L1 regularization: A comparative study and two new approaches. *18th Eur. Conf. Mach. Learn., ECML 2007* **2007**, *4701 LNAI*, 286–297.

(18) Schmidt, M.; Fung, G.; Rosales, R. *Optimization Methods for L1-Regularization*; Technical Report, 2009.

(19) Arngren, M.; Larsen, J.; Hansen, P. W.; Eriksen, B.; Larsen, R. *Analysis of Pre-Germinated Barley using Hyperspectral Image Analysis, Matlab Toolbox*; http://www.imm.dtu.dk/pubdb/p.php?5990; Technical Report, 2010.