



Near-infrared Spectroscopy and Hyperspectral Imaging for Sugar Content Evaluation in Potatoes over Multiple Growing Seasons

Ahmed M Rady^{1,2} · Daniel E Guyer³ · Nicholas J Watson¹

Received: 14 May 2020 / Accepted: 12 October 2020 / Published online: 28 October 2020
© The Author(s) 2020

Abstract

Sugar content is one of the most important properties of potato tubers as it directly affects their processing and the final product quality, especially for fried products. In this study, data obtained from spectroscopic (interactance and reflectance) and hyperspectral imaging systems were used individually or fused to develop non-cultivar nor growing season-specific regression and classification models for potato tubers based on glucose and sucrose concentration. Data was acquired over three growing seasons for two potato cultivars. The most influential wavelengths were selected from the imaging systems using interval partial least squares for regression and sequential forward selection for classification. Hyperspectral imaging showed the highest regression performance for glucose with a correlation coefficient (ratio of performance to deviation) or $r(\text{RPD})$ of 91.8(2.41) which increased to 94%(2.91) when the data was fused with the interactance data. The sucrose regression results had the highest accuracy using data obtained from the interactance system with $r(\text{RPD})$ values of 74.5%(1.40) that increased to 84.4%(1.82) when the data was fused with the reflectance data. Classification was performed to identify tubers with either high or low sugar content. Classification performance showed accuracy values as high as 95% for glucose and 80.1% for sucrose using hyperspectral imaging, with no noticeable improvement when data was fused from the other spectroscopic systems. When testing the robustness of the developed models over different seasons, it was found that the regression models had $r(\text{RPD})$ values of 55(1.19)–90.3%(2.34) for glucose and 35.8(1.07)–82.2%(1.29) for sucrose. Results obtained in this study demonstrate the feasibility of developing a rapid monitoring system using multispectral imaging and data fusion methods for online evaluation of potato sugar content.

Keywords Potatoes · Optical · Sugars · Evaluation · Multi-sensor data fusion · Machine learning

Introduction

Potato is a major root vegetable for human consumption and a known source of micronutrients including vitamins (C, niacin, B₆, and thiamine) and minerals (potassium, phosphorus, magnesium, and iron) (Rama and Narasimham 2003). Among several processed potato products, chips (sometimes called crisps) and French fries are the most common in both

developed and developing countries. The US per capita fresh tuber consumption decreased from 28.03 to 14.20 kg from 1970 to 2018. However, frozen potato and chip consumptions together increased from 20.82 to 31.53 kg per capita during the same period (NPC 2019). Moreover, the value of US exports of chips and frozen French fries increased from \$610 million in 2006 to more than \$1083 million in 2018 (Bohl and Johnson 2010; NPC 2019).

The quality and consistency of food products are important factors that often affect consumer purchasing decisions. Higher levels of sugars in potato tubers and consequently fried potato products can lead to the formation of harmful compounds and/or unfavorable color or taste (Stark and Love 2003). Among several quality attributes that attract consumers to fried potato products, color is the most important. During the frying process, a reaction takes place between the reducing sugars, mainly glucose and fructose in potatoes, and the amino acid, asparagine, at relatively high temperatures (around 180 °C). This phenomenon is known as the Maillard reaction

✉ Ahmed M Rady
radyahme@msu.edu

¹ Food, Water, Waste Research Group (FWW), Faculty of Engineering, University of Nottingham, University Park, Nottingham NG7 2RD, UK

² Department of Agricultural and Biosystems Engineering, Alexandria University, Alexandria, Egypt

³ Department of Biosystems and Agricultural Engineering, Michigan State University, East Lansing, MI 48824, USA

and results in a non-enzymatic browning color on the fried potato products (Stadler et al. 2002; Storey and Davies 1992). During frying, acrylamide is also formed which has toxic consequences for the nervous system and is a possible carcinogen factor in laboratory animals (Mottram et al. 2002; Stadler et al. 2002). Additionally, sweetening is not a desirable characteristic of fried potatoes and can result from high levels of sugars during storage (Stark and Love 2003). Therefore, monitoring sugar levels in potatoes used for frying is crucial for producing and preserving high-quality products that benefit growers, processors, and consumers. It is recommended that the optimal glucose level at harvest or during storage is 0.035% fresh weight (FW) for producing chips and 0.12% FW for French fries, whereas the desired sucrose levels for tubers dedicated for chips is 0.15% FW at harvest and 0.10% FW during storage. For tubers destined for French fries, the sucrose content should not exceed 0.15% FW during harvest or storage (Stark and Love 2003).

Measuring sugar levels in potatoes is usually conducted using laboratory-based methods including enzymatic hydrolysis, high-performance liquid chromatography (HPLC), high-performance anion-exchange chromatography (HPAEC), gas-liquid chromatography, and the YSI analyzer that was invented by Yellow Springs Instruments (Yellow Springs Instrument, Yellow Springs, OH, USA) (Rady and Guyer 2015c). Such techniques are usually destructive to samples, are laborious, and require skilled workers. Therefore, they are not suitable for real-time online inspection.

Near-infrared spectroscopy (NIR) is a rapid, relatively low-cost technique that has been used for quantitative and qualitative quality evaluation of agricultural products (Lohumi et al. 2015; Qu et al. 2015; Rady et al. 2020; Sahni et al. 2004). NIR diffuse reflectance spectroscopy can be used for detecting various chemical compounds depending on different light absorption mechanisms inside a material (Griffiths and Dahm 2007). Commercial NIR systems, especially those based on diffuse reflectance, have been successful in assessing the quality of fresh fruits and vegetables (Giovenzana et al. 2016) meat (Dixit et al. 2017), and grains (Singh et al. 2006), whereas hyperspectral imaging (HSI) is a method that combines spatial and spectral information about samples under inspection (Wu and Sun 2013). HSI systems have been used for monitoring the quality attributes of numerous agricultural products, including fruits (Ekramirad et al. 2017; Li et al. 2011; Li et al. 2018; Peng and Lu 2008), vegetables (Diezma et al. 2013; Huang et al. 2013), grains (Arngren et al. 2011; Serranti et al. 2013), and meats (Huang et al. 2013; Rady and Adedeji 2018).

Any sensor method requires the development of suitable models to relate recorded measurements to the sample properties of interest. Machine learning methods are one of the most popular modeling techniques as they are capable of processing large volumes of data and do not require the

development of complex physical models, which are difficult to develop using data from industrial environments. Machine learning methods can use supervised or unsupervised learning methods and can be used to develop regression and classification models. The combination of sensors and machine learning is ideally suited for the quality monitoring of food products as they can provide real-time information on the products once the models have been trained and can be applied in a variety of industrial environments (e.g., field and factory).

Several studies have been implemented for measuring quality attributes of potatoes using NIR techniques such as dry matter and specific gravity (Chen et al. 2005; Hartmann and Büning-Pfaue 1998; Helgerud et al. 2012; Subedi and Walsh 2009). The results of such studies showed correlation coefficient (r) values of 90–97% between predicted dry matter values using NIR and actual values determined using standard laboratory techniques. Measurement of sugar content in potatoes has also been studied using spectroscopic systems and hyperspectral imaging systems (Rady et al. 2015). The optimal r values of such studies were 14–98% whereas the value of the root mean square error of prediction (RMSEP) was as high as 0.89% FW. However, the previous studies developed models for individual cultivars and did not study sensor fusion.

According to Hall and Llinas (1997), sensor fusion “combines data from multiple sensors, and related information from associated databases, to achieve improved accuracies and more specific interferences than could be achieved by the use of a single sensor alone.” The data combined from each sensor should, however, provide distinguishing and non-redundant information about the measured property for a benefit to be observed in the regression and/or classification models used to analyze the sensor data. Data fusion is conducted by either concatenating the features from various sensors, then processing them, or performing feature selection before combining and processing (Manso 2008). Fusing data that were acquired from different electronic sensors has been studied for improving the regression and/or classification models of quality attributes of fruits, vegetables, and other food products (Manso 2008). The fusion of data obtained by stationary and online prototype hyperspectral imaging systems was conducted to improve the regression capability of firmness and soluble solid content (SSC) for three apple cultivars and the standard error of prediction (SEP) values decreased by 16.1% for firmness, and 11.2% for SSC (Mendoza et al. 2011). Another study was conducted on the same apple cultivars to examine the fusion of data obtained from spectroscopic, hyperspectral imaging, acoustic firmness, and bioyield firmness to assess apple firmness and the SSC and SEP values decreased by 20% for firmness and 6% for SSC when using sensor fusion (Mendoza et al. 2012). Additional examples include combining data from an electronic tongue (e-tongue) and a spectroscopic system for determining the botanical

origin of honey (Ulloa et al. 2013), and fusing ultraviolet (UV)/Vis, Vis/NIR, capacitance and conductance, ultrasonic, and color sensors (Ignat et al. 2014) for predicting several maturity indices for bell peppers.

This study aims to investigate the use of optical systems, data fusion at the feature level, and machine learning algorithms to evaluate the quality of potato tubers based on their glucose and sucrose content. Quality assessment models were developed that do not depend on the tuber variety or growing season. A range of different regression and classification machine learning models were developed utilizing data from individual spectroscopic systems in addition to data fused from a combination of systems. Two different potato cultivars were studied to ensure models were developed that were not cultivar-specific. Measurements were performed over three growing seasons to assess the performance of models developed from one season on unseen data on another season.

The combination of optical measurements and machine learning models will enable rapid and nondestructive determination of sugar content within stored potatoes enabling appropriate storage strategies and ultimately improve the quality of potato products. Testing the models with data from different seasons will significantly increase the potential of the techniques as they will not require new models to be developed each year. While regression models provide quantitative estimates of the sugar levels in potatoes, classification models also benefit storage technicians by quickly indicating if the sugar level is within an acceptable range. This information can be used to determine if the potatoes are suitable for processing or storage conditions (e.g., temperature) require modification.

Materials and Methods

Raw Materials and Sampling

In this study, the cultivars of Frito Lay 1879 (FL, commonly used for frying) and Russet Norkotah (RN, commonly used

for baking) were utilized. These potatoes were grown, stored, and measured over the 2008, 2009, and 2011 growing seasons. Table 1 summarizes the different sources, storage conditions, experiment periods, number of tested tubers, and the electronic sensor systems used in the studied seasons. In each season, tubers were first cleaned, and any bruised or externally damaged tubers were discarded before storage and sampling. In the 2008 season, tubers were collected from a commercial farm in Southwest Michigan, USA, and stored at either 7, 10, or 15 °C. During this season, a total of 200 samples (i.e., tubers) of each cultivar were tested over 130 days at 4 sampling times. In the 2009 season, samples were collected from two different locations: a research farm at Montcalm County, MI, USA, and the Michigan State University Muck experimental farm, Bath, MI, USA. Tubers were then stored at 4, 7, and 10 °C and were sampled monthly from November 2009 to April 2010. The total number of tubers measured was 540 for FL and 180 for RN. In the 2011 season, samples were brought from a commercial farm in Southwest Michigan, USA, and stored at 1, 4, 7, 10, and 13 °C. Sampling took place monthly starting November 2011 until May 2012 and included 195 and 75 tubers for FL and RN, respectively. The variation of samples' sources and storage temperatures were implemented to create a broad range of sugar concentrations in the tubers. This was required to ensure the developed machine learning models were robust and could be used on different cultivars from different seasons, where large variations in sugar content may exist. Three optical systems were utilized to acquire spectral information from the whole tubers. These systems were as follows: (1) Vis/NIR interactance, (2) Vis/NIR hyperspectral imaging, and (3) NIR reflectance. During the 2008 season, Vis/NIR interactance in addition to Vis/NIR hyperspectral imaging systems were used, whereas, in the 2009 and 2011 seasons, Vis/NIR interactance and NIR reflectance systems were utilized. It was not possible to use all three spectroscopic systems every year due to equipment availability. More details for the studies conducted on the various seasons can be obtained from previous studies (Rady and Guyer 2015b; Rady et al. 2014). It should be stated that all previous studies were

Table 1 Various configurations of the experiments of testing Frito Lay 1879 and Russet Norkotah potato cultivars using spectroscopic and hyperspectral imaging systems

	2008 season	2009 season	2011 season
Raw material source	A commercial farm in Southwest Michigan, USA	Two research farms in Montcalm and Bath, MI, USA	A commercial farm in Southwest Michigan, USA
Storage conditions	7, 10, and 15 °C, RH = 90–95%	4, 7, and 10 °C, RH = 90–95%	1, 4, 7, 10, and 13 °C, RH = 90–95%
Experiments time span	September, 2008–February, 2009	November, 2009–April, 2010	November, 2011–May, 2012
Number of tested tubers	200 for FL, 200 for RN	540 for FL, 180 for RN	195 for FL, 75 for RN
Electronic systems used for scanning	VIS/NIR interactance, VIS/NIR hyperspectral imaging	VIS/NIR interactance, NIR reflectance	VIS/NIR interactance, NIR reflectance

based on producing models for each cultivar and/or using optical systems individually without investigating data fusion. Figure 1a–c show a schematic diagram of the different electronic systems used in this study, whereas Fig. 1d–f shows examples of the output signals obtained from these systems.

Vis/NIR Interactance System

The Vis/NIR interactance system records measurements on a tuber which is placed directly on the integrated fiber optic probe as shown in Fig. 1a. Both incident and reflected light are vertical to the sample under inspection (Rady and Guyer 2015a). The system included a spectrometer (model no. USB 4000, Ocean Optics, Inc., Dunedin, FL, USA), fiber optic (200- μm diameter), and a light source with a maximum power of 250 W (model no. 66881, Oriel Inst., Irvine, CA, USA). The system captures the reflected light in the range of 446 to 1125 nm with a Full-Width Half Maximum (FWHM) of 0.3 nm and an integration time of 10 ms. Each acquired spectrum such as the one in Fig. 1d was normalized using the signal acquired from a Teflon® disc and the relative interactance was then calculated (Rady and Guyer 2015a).

Vis/NIR Hyperspectral Imaging System

The hyperspectral imaging (HSI) system produced back-scattered images (256×256 pixels) in the wavelength range of 400–1000 nm with spatial and spectral resolutions of 0.2 mm/pixel and 2.35 nm, respectively. The system as shown in Fig. 1b contained a CCD camera (model no. C4880, Hamamatsu Photonics, Hamamatsu, Japan), spectrograph (ImSpector V10, Spectral Imaging Ltd., Oulu, Finland), power supply (model no. 69931, Oriel Inst., Irvine, CA, USA), exposure controller (model no. 68945, Oriel Inst., Irvine, CA, USA), DC power supply (model no. 65423A, Agilent Tech., Santa Clara, CA, USA). Each image, as shown in Fig. 1e, was then normalized using a Teflon® reference cube, and the Mean Reflectance Spectra (MRS) was extracted (Rady et al. 2015).

NIR Reflectance System

The NIR reflectance system utilized in this study (Fig. 1c) had an InGaAs spectrometer (model no. NIR512L-1.7T1, Control Development, Inc., South Bend, In, USA), power supply

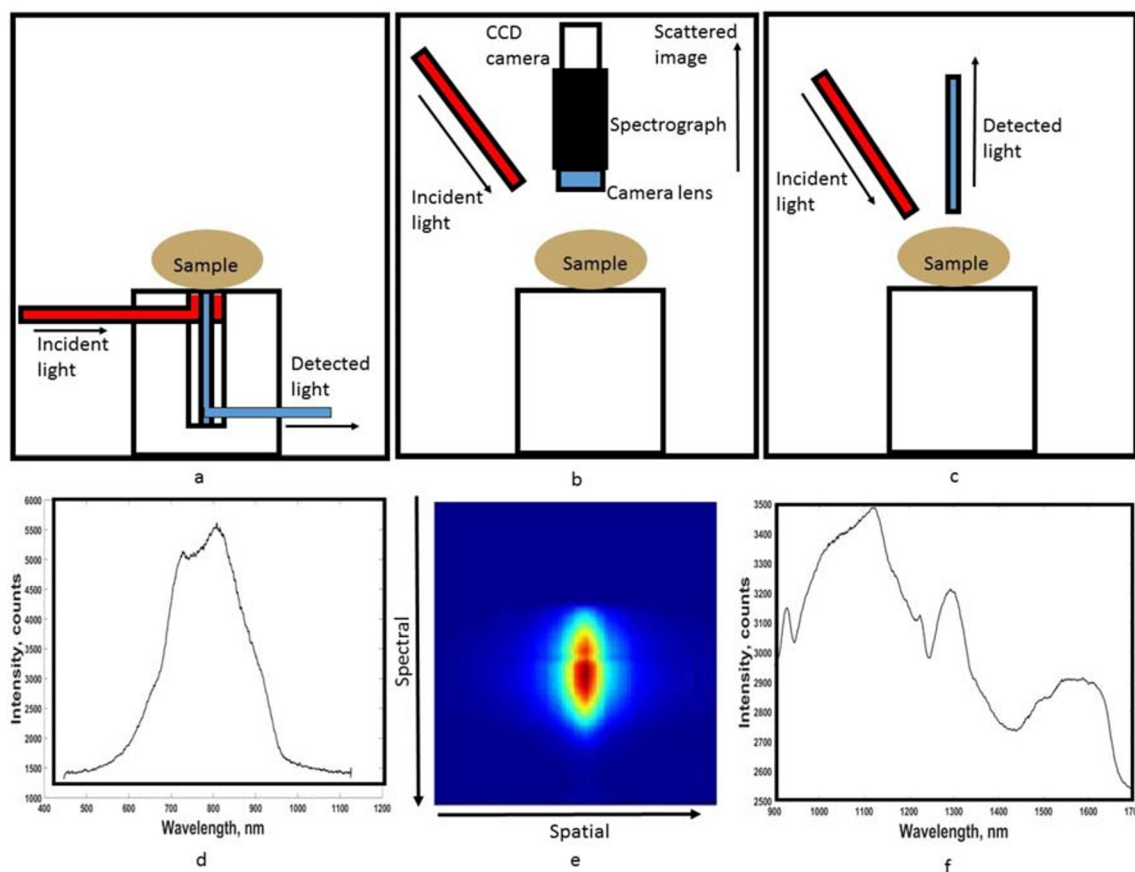


Fig. 1 Different optical systems along with typical recorded spectra/image. **a, d** Interactance spectroscopy. **b, e** Hyperspectral imaging. **c, f** Reflectance spectroscopy

(model no. 68931, Oriel Inst., Irvine, CA, USA) with a maximum power of 300 W, light source (model no. 66881, Oriel Inst., Irvine, CA, USA). The system operates in the diffuse reflectance mode in the NIR range (900–1685 nm). The system has an FWHM value of 3.25 nm and the integration time was 8 ms. The obtained spectrum for each sample, as shown in Fig. 1f, was normalized relative to measurements from a Teflon® disc, and then the relative reflectance was calculated (Rady and Guyer 2015b).

Wet Chemistry Determination of Sugar Content

To determine the glucose and sucrose concentrations for each sample, an enzymatic technique was performed using the Megazyme sucrose/D-glucose assay procedure (Megazyme International Ireland Ltd., Wicklow, Ireland) (Rady et al. 2014). After performing the electronic measurements, samples were stored in labeled plastic bags and placed in an ice-containing foam box until the juice was then extracted from each tuber. A Juicerator (ACME Supreme, New Hartford, CT., USA) was used to extract the juice from tubers, and the extracted juice for each sample was transferred into a polystyrene tube and stored at -20°C for later analysis. The procedure presented in Rady et al. (2014) was followed to obtain the glucose and sucrose concentrations (gram per 100-g fresh tuber weight glucose % and sucrose %).

Machine Learning Models for Classification and Prediction of Sugar Content in Potato Tubers

Wavelength Selection for Regression, Classification, and Data Fusion

The main objectives of wavelength selection are to overcome the overfitting problem associated with high dimensional data, reduce computational time, and consequently improve regression and/or classification performance (Heise and Winzenm 2001; Mark 2001; Varmuza and Filzmoser 2016). Data dimensional reduction is particularly important for the HSI due to large volumes of data these systems produce. This is especially the case if they are to be used to develop multispectral imaging systems for online inspection systems where the time available to measure each sample is limited. In this study, the interval partial least squares (IPLS) method was applied for wavelength selection for the regression tasks. The IPLS configuration used in this study included the forward mode, a window width (i.e., the number of variables selected at each run of the partial least squares) of 1, 2, or 3 variables, and 20 latent variables. For the classification models, the sequential forward selection (SFS) technique was used for wavelength selection. The SFS method depends on starting with an empty model (i.e., no selected variables or features) and adding a new feature each time before evaluating the model with the

appropriate classification technique (partial least squares discriminant analysis (PLS-DA) in this work). Based on the performance of the resulting model, the feature will be kept if the performance of the current model is higher than that of the previous model or, otherwise discarded (Mao 2004).

After obtaining the selected wavelengths, regression (using calibration then validation), or classification (using training then testing) models were built for the individual spectroscopic systems. Following this, the data from the different systems were fused such that in the 2008 season, the data obtained from hyperspectral and interactance systems were concatenated, whereas, in the 2009 and 2011 seasons, the data obtained from Vis/NIR interactance and NIR reflectance systems were concatenated. To develop regression and classification models that included data over multiple seasons, interactance data from 2008, 2009, and 2011 were combined, and the latter two season's reflectance data were also combined. Additionally, to test the robustness of the data over different seasons models developed in one season were tested with unseen data from the other seasons. For example, interactance data obtained from the 2008 season was used as calibration or training whereas the 2009 or 2011 data were used for regression or testing. It should be noted that the data was normalized before building regression or classification models. A visual representation of the sensor fusion process is shown in Fig. 2.

Data Preprocessing

Several preprocessing methods were applied to the data to minimize spectral noise, baseline shifts, and to overcome variation of sample condition due to temperature and natural variability (Christy and Kvalheim 2007). The preprocessing methods included, mean centering, first derivative smoothing, second derivative smoothing, standard normal variate (SNV) correction, and multiplicative signal correction (MSC). Besides, a logarithmic transformation was applied to the raw sugar content data to reduce any skewness in the original distribution which may affect the performance of the regression and classification models (Varmuza and Filzmoser 2016).

Partial Least Squares Regression for Sugar Prediction

To build the regression models for sucrose and glucose content in the tubers, partial least squares regression (PLSR) was utilized. This was selected as it is an effective linear regression technique capable of processing collinear high dimensional data (Varmuza and Filzmoser 2016). The number of latent variables was chosen as 20 based on preliminary work and previous studies (Rady and Guyer 2015b). For each regression model, the data was divided into calibration (80%) and validation (20%). Cross-validation (4-fold) was applied to the data to obtain the best calibration model based on the

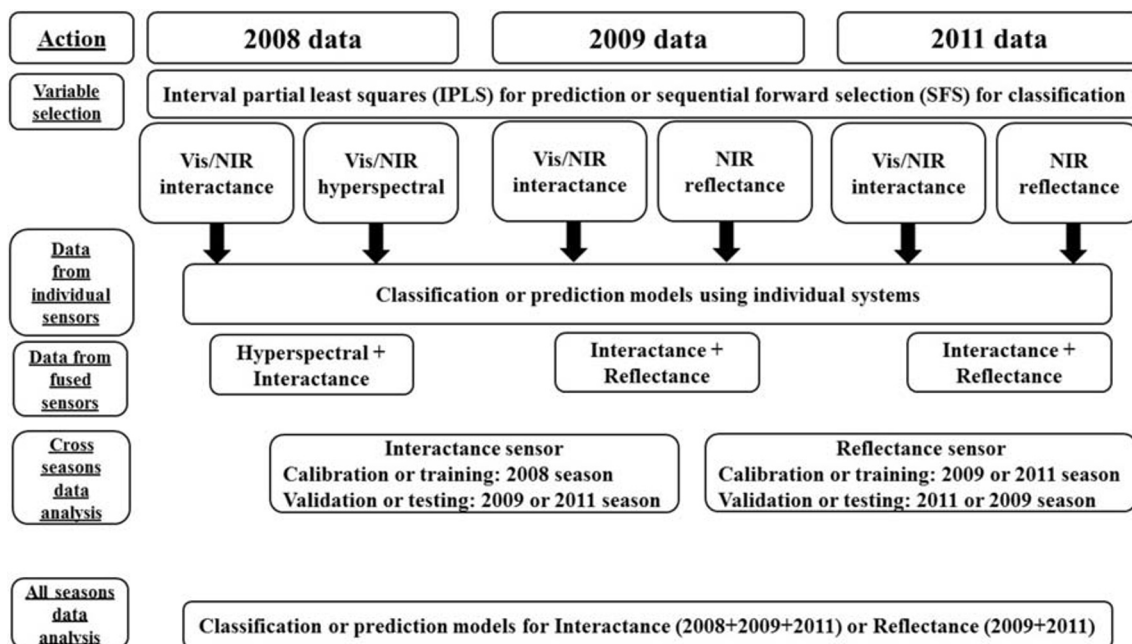


Fig. 2 Workflow for processing the data obtained from different seasons for evaluation of glucose and sucrose in potato tubers

minimum root mean square error of calibration using cross-validation (RMSEC), correlation coefficient (r), and the ratio between the standard deviation of the reference data (i.e., glucose or sucrose), and the RMSEC (RPD). PLSR algorithm was implemented in this study using the MATLAB® statistical toolbox.

Classification of Potato Tubers Based on Sugar Levels

In this study, several classification techniques were studied for the individual spectroscopic sensors and fused sensor data. The methods included linear discriminant analysis (LDA), K-nearest neighbor (Knn), PLS-DA, and artificial neural networks (ANN). In the case of LDA, the Euclidean distance was used to assign each sample to a certain class, and principle component analysis (PCA) was conducted on the fused spectral data to avoid the collinearity problem (Rady et al. 2020). The components responsible for > 99% of the total variation between tubers were considered in the subsequent classification tasks (Duda et al. 2012). In the case of Knn, the Euclidean distance with k values of 4 was selected (Rady et al. 2020). For the PLS-DA, 20 latent variables were used to build classification models (Rady and Guyer 2015b). Finally, the ANN was a feed-forward neural network that contained an input layer of the pretreated data, a hidden layer of 50 neurons, and an output layer with the assigned classes (high/low sucrose/glucose content). Log-sigmoid was chosen as the transfer function, and the scaled conjugate gradient backpropagation method for hidden and output layers, respectively (Rady et al. 2020). All specific values for parameters for different classifiers are based on preliminary try and error analysis or previous studies.

Data dedicated for classification were preprocessed as stated before in the case of regression. To build classification models, the data was divided into a training set (80%) and a testing set (20%). Training data was then used to build classification models using 4-fold cross-validation to enhance model robustness. Spectral and sugar data in each season were divided into two classes based on sugar concentrations, using cut-off values that were adopted from the literature (Stark and Love 2003). The cut-off values for glucose or sucrose were chosen as the median for each season's data. Table 2 shows the cut-off as well as the minimum and maximum values for glucose and sucrose for each season. It is worth stating that cut-off values are within the values recommended by previous work (Stark and Love 2003). The classification models developed using LDA and ANN were conducted using the MATLAB® statistical toolbox, whereas the models developed using PLS-DA and Knn used the MATLAB toolbox by Davide Ballabio (Milano Chemometrics and QSAR Research Group, University of Milano-Bicocca, Milan, Italy) (Ballabio and Consonni 2013).

Results and Discussion

Wavelengths Selected for Regression and Classification Models

The selected wavelengths for glucose or sucrose in potato tubers obtained for all sensors are shown in Fig. 3 for regression using IPLS and Fig. 4 for classification using SFS. Regression is based on studying the correlation or closeness

Table 2 Cut-off value (minimum-maximum, standard error values) for glucose and sucrose for the 2008, 2009, and 2011 seasons

Constituent	2008	2009	2011
Glucose	0.0557 (0.00–4–2.1458, 0.4513)	0.0366 (0.0028–0.3574, 0.0545)	0.0847 (0.0224–1.1355, 0.2231)
Sucrose	0.0839 (0.00–4–2.5439, 0.2836)	0.0393 (0.00–5–0.4205, 0.0750)	0.0788 (0.00–2.1842, 0.2166)

between the predicted vs the measured values of a certain constituent, then the efficacy will be estimated based on a value to value basis, whereas in the case of classification, the objects might be clustered to the same class even if they have different values. Therefore, there is a significant increase in the number of selected wavelengths selected from all sensors for the regression models (Figs. 3 and 4). This highlights one of the benefits of classification models when compared to regression. Although they do not provide quantitative information on a sample (e.g., sugar content), they require less data to create models and can, therefore, be developed using cheaper hardware, making them more suitable for industrial applications.

In the case of regression, more wavelengths were selected for glucose prediction models than those for sucrose. Hyperspectral imaging showed 70 selected wavelengths in 2008 whereas interactance data showed 77, 61, and 98 in 2008, 2009, and 2011, respectively. While the hyperspectral imaging resulted in selected wavelengths for glucose located in both the visible (400–780 nm) and NIR (780–2500 nm)

wavelength ranges, the interactance data showed all selected wavelengths only in the visible range for the three seasons. For the NIR reflectance data, a considerable increase of selected wavelengths was observed in the 2009 (84) and 2011 (106) seasons compared with the interactance data. In the case of sucrose, the selected wavelength from the hyperspectral imaging was 19 with the majority (17) selected in the visible range, whereas those from the interactance were 61, 8, and 34 in the 2008, 2009, and 2011 seasons, respectively, and all located in the visible range. Reflectance data, however, resulted in a lower number of selected wavelengths in the 2009 (35) and 2011 seasons (13) than the other spectroscopic systems.

For the classification models, the number of selected wavelengths for glucose from hyperspectral imaging was 9 with 3 in the NIR region, whereas 11, 4, and 8 wavelengths were selected for the interactance data in the 2008, 2009, and 2011 seasons, respectively, with all being located in the visible range. For the NIR reflectance data, only 8 in 2009 and 3 in 2011 were selected. For sucrose, the hyperspectral system resulted in 11 selected wavelengths in 2008 compared with 3,

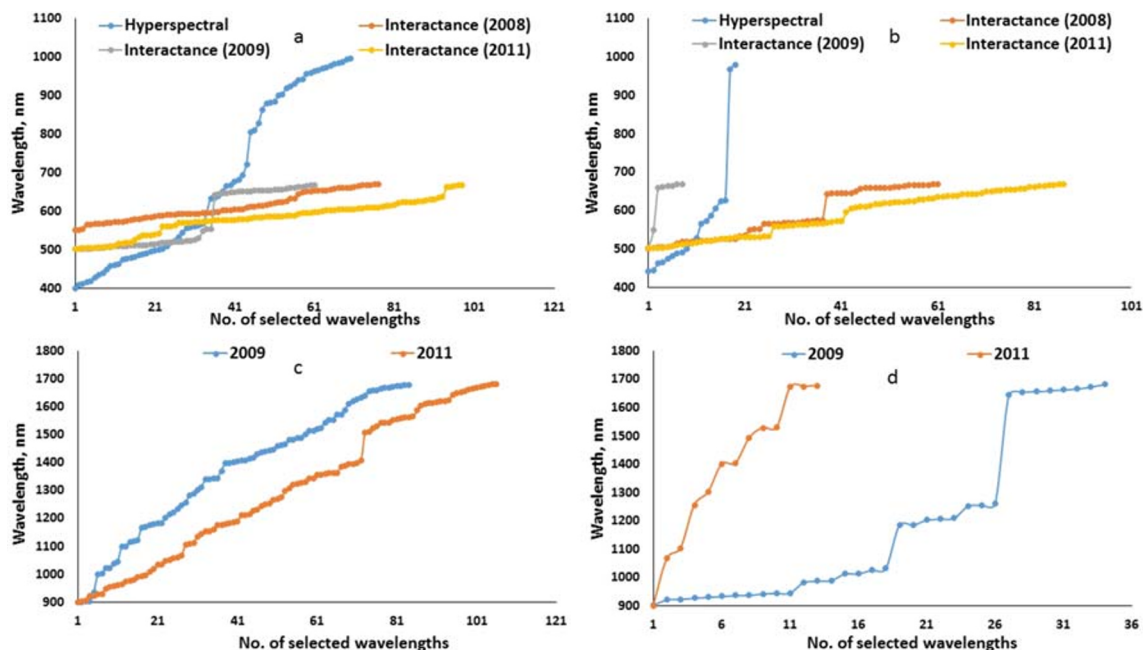


Fig. 3 Selected wavelengths for predicting glucose or sucrose levels in potato tubers on obtained from the 2008, 2009, and/or 2011 seasons for **a** glucose, hyperspectral imaging, and interactance; **b** sucrose, hyperspectral imaging, and interactance; **c** glucose and reflectance; and **d** sucrose and reflectance

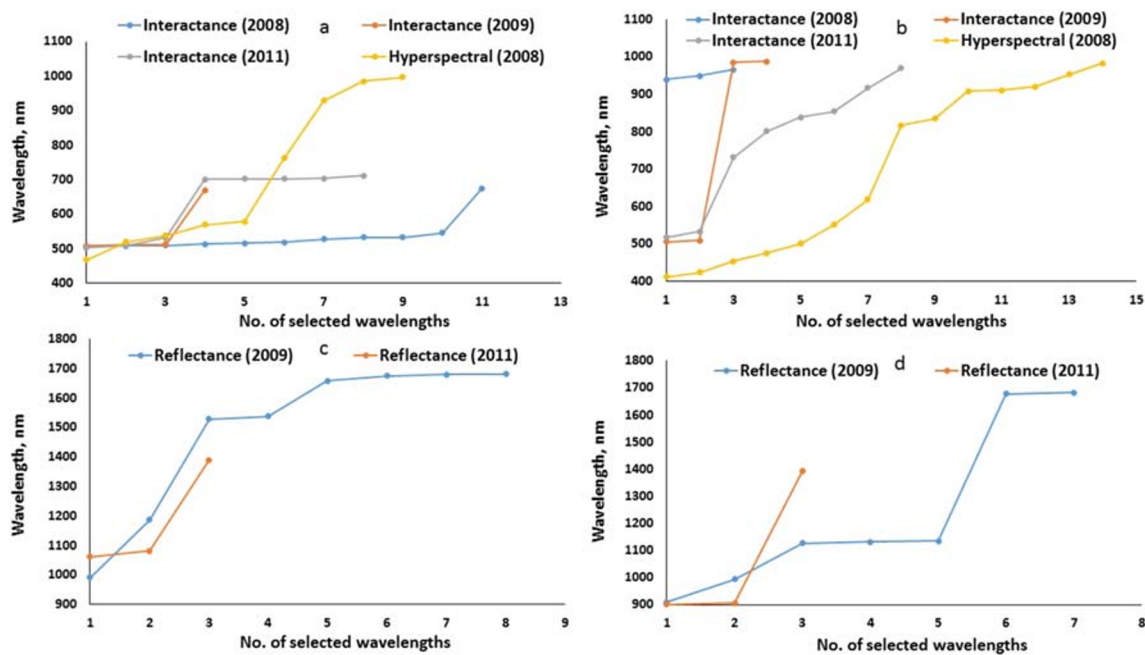


Fig. 4 Selected wavelengths for classifying potato tubers based on glucose or sucrose levels obtained from the 2008, 2009, and/or 2011 seasons for **a** glucose, hyperspectral imaging, and interactance; **b** sucrose, hyperspectral imaging, and interactance; **c** glucose and reflectance; **d** sucrose and reflectance

4, and 8 for the interactance system in the 2008, 2009, and 2011 seasons, respectively. The selected wavelengths for sucrose prediction from both sensors contained NIR information in contrast to the glucose case stated earlier where only the hyperspectral imaging system contained NIR region selected wavelengths. Only 7 and 3 wavelengths were selected for the 2009 and 2011 seasons, respectively. Considering the different storage temperatures and different growing conditions (i.e., soil and weather) and having two different cultivars, it is expected that the selected wavelengths in one season might not match those in other seasons. However, the selected wavelengths in this study contain those identified by other researchers studying food materials. A wavelength of approximately 913 nm was associated with sugar absorption in potatoes by previous researchers (Yaptenco et al. 2000). Moreover, the wavelengths of 1190 and 1400 nm were identified as key variables in assessing quality parameters during the ripening of four grape cultivars (Kemps et al. 2010). These wavelengths were associated with the presence of the O-H group linked to sugar absorption in grapes. Soluble solids which are an indication of sugar content were stated to be related to the C-H band at 910 nm and the O-H band at 950 nm in grape, lime, and star fruit (Fairuz Omar 2013), and 960 nm for the O-H group in navel orange (Liu et al. 2010). In addition to 960, 1180 and 1450 nm have also been associated to the O-H group in jujube fruit (Zhang et al. 2012), and 975 due to the O-H group in blueberry (Shao et al. 2006). Generally, the wavelengths of 972 and 1009 nm are related to the second overtone of the O-H group which is related to the

presence of saccharides (Workman and Weyer 2012). In this current work, most of the aforementioned values were present in the selected wavelengths used to develop the regression and classification models (Figs. 3 and 4). However, there is no exact match of the selected wavelengths between different fruits and vegetables which is mostly a result of the various concentration of sugars and the different compositions among different fresh products.

Partial Least Squares Regression for Sugar Content Prediction

The cross-validation results for the regression of glucose and sucrose content obtained from individual sensors (IS) as well as fused data (FD) are shown in Table 3. The best models for each highlighted in Tables 3, 4, 5, and 6. The highest performing glucose content prediction models were obtained from the hyperspectral imaging system in 2008 with an $r(\text{RPD})$ value of 91.8%(2.41), whereas in 2009 and 2011, either the interactance or reflectance models showed similar regression performance with slightly better performance for the reflectance system with $r(\text{RPD})$ values of 60%(1.04). In the case of the fused data, the performance of the PLSR models significantly improved, especially for the 2009 and 2011 seasons. Combined hyperspectral and interactance data for the 2008 season resulted in $r(\text{RPD})$ values of 94%(2.91). In the case of interactance and reflectance fused data, $r(\text{RPD})$ values increased to 68.2%(1.37) and 83.6%(1.86) for the 2009 and 2011 seasons, respectively. Models obtained from

Table 3 Best PLSR prediction models for glucose and sucrose of potato tubers using: Vis/NIR interactance, Vis/NIR hyperspectral imaging, and NIR reflectance systems obtained from selected wavelengths for data combined from Frito Lay 1879 and Russet Norkotah cultivars in the 2008, 2009, and 2011 seasons. Rows with values in boldface represent the best models for each season

Constituent	Season	Sensor	Preprocessing method (spectra, constituent)*	Cross-validation performance*		
				r_{cv} (%)	RMSE _{cv} (%)	RPD _{cv}
Glucose	2008	VIS/NIR hyperspectral imaging	MC, Non	91.8	0.0204	2.41
		VIS/NIR interactance	MSC, Non	69.7	0.0335	1.38
	2009	VIS/NIR interactance	MC, Non	47.7	0.0186	1.14
		NIR reflectance	MC, Non	43.4	0.0199	1.10
	2011	VIS/NIR Interactance	MC, Non	54.1	0.0389	1.20
		NIR reflectance	MS, Non	60	0.0332	1.04
	2008	VIS/NIR hyperspectral imaging + VIS/NIR interactance	MC, Non	94	0.0182	2.91
		VIS/NIR interactance + NIR reflectance	MSC, Non	68.2	0.0164	1.37
	2011	VIS/NIR interactance + NIR reflectance	2 nd Derv, Non	83.6	0.0222	1.86
	2008–2011	VIS/NIR interactance	MSC, Power	65.4	0.0555	1.32
2008–2011	NIR reflectance	SNV, Non	62.6	0.0219	1.27	
Sucrose	2008	VIS/NIR hyperspectral imaging	MC, Non	61.2	0.0243	1.26
		VIS/NIR interactance	MC, Non	50.7	0.0314	1.15
	2009	VIS/NIR interactance	MC, Non,	74.5	0.0187	1.40
		NIR reflectance	MC, Non	65.6	0.0186	1.32
	2011	VIS/NIR interactance	2 nd Derv, Non	56.1	0.0362	1.22
		NIR reflectance	MC, Power	60.7	0.0634	1.21
	2008	VIS/NIR hyperspectral imaging + VIS/NIR interactance	MSC, Non	84.2	0.0239	1.37
		VIS/NIR interactance + NIR reflectance	MSC, Non	71.4	0.0180	1.40
	2011	VIS/NIR interactance + NIR reflectance	MC, Non	84.4	0.0354	1.82
	2008–2011	VIS/NIR interactance	SNV, Power	62.8	0.0608	1.28
2008–2011	NIR reflectance	2 nd Derv, Non	58.0	0.0276	1.09	

* r_{cv} , coefficient of correlation using cross-validation; RMSE_{cv}, root mean square error of calibration using cross-validation; RPD_{cv}, ratio of performance to deviation using cross validation

** MC, mean centering; Non, no preprocessing; MSC, multiplicative signal correction; 2nd Derv, smoothing using the second derivative

individual systems for sucrose content prediction resulted in the highest r (RPD) value of 61.2%(1.26) in the 2008 season using hyperspectral imaging, 74.5%(1.40) in the 2009 season using the interactance system, and 60.7%(1.21) in the 2011 season using the reflectance system. Sucrose prediction models also showed considerable improvement after using data fusion. Fused data in the 2008 season resulted in r (RPD) values of 84.2%(1.37) and 84.4%(1.82) in the 2011 season, whereas in 2009, the r (RPD) values were slightly lower than those deduced from the individual systems. Regression models of glucose content showed that the hyperspectral imaging system produced the best performance. However, no improvement was achieved using fused data for the regression of sucrose content. Regression model results using all seasons' data for the interactance system showed r (RPD) values of 65.4%(1.32) for glucose and 62.8%(1.28) for sucrose, whereas the values obtained from reflectance data were 62.6%(1.27) for glucose and 58%(1.09) for sucrose. The

regression results can be attributed to the data variation among different seasons highlighting the challenge of developing multi-season models.

The results obtained from the models developed in this current study had lower r (RPD) values compared with those from previous research, based on single cultivars (FL or RN) and individuals sensors (IS) (Rady and Guyer 2015a; Rady et al. 2015; Rady and Guyer 2015b; Rady et al. 2014) The r (RPD) values determined from the earlier studies for the optimal glucose concentration evaluation models were 88%(1.78) and 97%(4.16) for FL and RN, respectively. Similarly, the optimal IS-sucrose models in earlier studies resulted in r (RPD) values of 88%(1.64) and 94%(2.82) for FL and RN. Other studies that tested whole tubers with optical systems and machine learning had r (RMSEP) values of 83%(0.087) for glucose and 95%(0.341) for sucrose (Yaptenco et al. 2000), and 65%(0.046) for glucose (Chen et al. 2010). In another

Table 4 Best classification models for potato tubers based on glucose and sucrose of potato tubers using individual and fused data from: Vis/NIR interactance, Vis/NIR hyperspectral imaging, and NIR reflectance systems obtained from selected wavelengths for data combined from Frito Lay 1879 and Russet Norkotah cultivars in the different seasons. Rows with values in boldface represent the best models for each season

Constituent	Season	Sensor	Preprocessing method (spectra, constituent)*	Classification technique	Classification accuracy (%)	
					Training set	Testing set
Glucose	2008	VIS/NIR hyperspectral imaging	SNV, Log	Knn	95.0	91.3
		VIS/NIR interactance	SNV, Log	Knn	91.4	87.5
	2009	VIS/NIR interactance	Non, Log	Knn	66.1	62.6
		NIR reflectance	MSC, Log	Knn	64.6	60.6
	2011	VIS/NIR interactance	MC, Log	PLS-DA	90.9	76.9
		NIR reflectance	SNV, Log	Knn	75.0	75.0
	2008	VIS/NIR hyperspectral imaging + VIS/NIR interactance	SNV, Non	Knn	93.8	92.5
	2009	VIS/NIR interactance + NIR reflectance	Non, Log,	PLS-DA	69.6	65.7
	2011	VIS/NIR interactance + NIR reflectance	Non, Log,	PLS-DA	81.8	78.8
	2008–2011	VIS/NIR interactance	MSC, Log	Knn	75.3	74.7
	2008–2011	NIR reflectance	SNV, Log	Knn	75.4	72.7
	Sucrose	2008	VIS/NIR hyperspectral imaging	SNV, Log	Knn	80.1
VIS/NIR interactance			SNV, Log	PLS-DA	75.5	65.0
2009		VIS/NIR interactance	1st Derv, Log	Knn	71.4	70.7
		NIR reflectance	SNV, Log	Knn	71.6	67.1
2011		VIS/NIR interactance	MC, Log	Knn	73.8	69.2
		NIR reflectance	MSC, Log	PLS-DA	66.9	61.5
2008		VIS/NIR hyperspectral imaging + VIS/NIR interactance	SNV, Non	Knn	80.5	77.5
2009		VIS/NIR interactance + NIR reflectance	MSC, Log	PLS-DA	73.9	73.6
2011		VIS/NIR interactance + NIR reflectance	Non, Log	PLS-DA	76.8	75.0
2008–2011		VIS/NIR interactance	SNV, Log	Knn	68.9	66.7
2008–2011		NIR reflectance	MSC, Log	Knn	75.5	71.9

* SNV, standard normal variate; Log, logarithmic; MSC, multiplicative signal correction; MC, mean centering; 1st Derv, smoothing using the first derivative; Non, no preprocessing

study, reducing sugars were assessed using Fourier transform near-infrared (FT-NIR) (800–2500 nm) on three potato cultivars and PLSR models yielded determination coefficient, R^2 values of 63–84% for glucose concentration (Camps and Camps 2019). The regression results yielded in this study by fusing data from different sensors are comparable to the listed results with the advantage that the models developed in this current work are not cultivar specific.

Classification of Potato Tubers Based on Sugar Content

The highest performing classification models of the studied tubers based on glucose and sucrose content and using data

acquired from the individual as well as fused sensors are shown in Table 4. In general, the Knn and PLS-DA methods resulted in the best classification performance. Classification models developed using data from the hyperspectral imaging or interactance systems showed similar performance in the 2008 season. In the 2009 and 2011 seasons, models developed from interactance or reflectance sensors also showed similar performance. The highest classification accuracy values for the testing set for the 2008, 2009, and 2011 seasons were 91.3, 62.6, and 76.9%, respectively, for the glucose and 78.8, 70.7, and 69.2% for sucrose. In the case of fused data, classification was generally enhanced for all seasons. The values of classification accuracy for fused data models were 92.5, 65.7, and 78.8% for the 2008, 2009, and 2011 seasons, respectively, in the case of

Table 5 Best PLSR prediction models for glucose and sucrose of potato tubers using: Vis/NIR interactance, Vis/NIR hyperspectral imaging, and NIR reflectance systems obtained from selected wavelengths for data combined from Frito Lay 1879 and Russet Norkotah cultivars based on testing the data over different seasons. Rows with values in boldface represent the best models for each season

Constituent	Season (calibration)	Season (validation)	Sensor	Preprocessing method (spectra, constituent)*	Calibration**			Prediction using cross-validation		
					r_c (%)	RMSE _c (%)	RPD _c	r_{cv} (%)	RMSE _{cv} (%)	RPD _{cv}
Glucose	2008	2009	VIS/NIR interactance	SNV, Non	96.5	0.0157	2.89	90.3	0.0227	2.34
	2008	2011	VIS/NIR interactance	MC, Non	94.3	0.0159	3.12	92.1	0.0205	2.59
	2009	2011	NIR reflectance	MC, Non	75.6	0.0247	2.15	68.6	0.0356	1.38
	2011	2009	NIR reflectance	MC, Non	64.1	0.0197	1.89	55.0	0.0202	1.19
	2009	2011	VIS/NIR interactance + NIR reflectance	2 nd Derv, Non	89.4	0.0208	1.97	83.6	0.0222	1.86
	2011	2009	VIS/NIR interactance + NIR reflectance	MSC, Non	76.4	0.0136	1.75	68.2	0.0164	1.37
Sucrose	2008	2009	VIS/NIR interactance	MSC, Non	87.4	0.0178	1.65	82.2	0.0255	1.29
	2008	2011	VIS/NIR interactance	MC, Non	83.7	0.0198	1.56	77.7	0.0222	1.47
	2009	2011	NIR reflectance	MC, Non	60.1	0.0425	1.45	42.4	0.0512	1.07
	2011	2009	NIR reflectance	MSC, Non	44.8	0.0176	1.57	35.8	0.0231	1.07
	2009	2011	VIS/NIR interactance + NIR reflectance	MSC, Non	89.4	0.0289	1.86	80.8	0.0396	1.66
	2011	2009	VIS/NIR interactance + NIR reflectance	MSC, Non	78.4	0.064	1.57	71.4	0.0180	1.40

* SNV, standard normal variate; Non, no preprocessing; MC, mean centering; 2nd Derv, smoothing using the second derivative; MSC, multiplicative signal correction

** r_c , coefficient of correlation for calibration; RMSE_c, root mean square error for calibration; RPD_c, ratio between the standard deviation of the constituent to the RMSE_c; r_{cv} , coefficient of correlation for prediction using cross-validation; RMSE_{cv}, root mean square error for prediction using cross-validation; RPD_{cv}, ratio between the standard deviation of the constituent to the RMSE_{cv}

glucose. In the case of sucrose, classification accuracy did not improve with data fusion for the 2008 seasons (77.5%), whereas an improvement was achieved for the 2009 (73.6%) and 2011 (75%) seasons. Results of classification based on glucose for season-combined data represented accuracy values of 74.7% for interactance, and 72.7% for reflectance. While those based on sucrose models were 66.7 for interactance and 71.9% for reflectance. Such results are comparable to the performance obtained from a single season for interactance or reflectance sensors which shows that data variation between seasons did not lead to lower performance models as indicated in the regression task. It was obvious that the models developed from the hyperspectral data resulted in much better classification for the 2008 season compared with the interactance data, which was similar to the regression results. Interactance and reflectance sensors showed similar performance, especially for the 2011 season. The Knn and PLS-DA along with SNV and non-preprocessed spectral data showed the optimal classification performance among the applied classifiers. In the case of glucose, fused data were found to slightly improve the classification results by 1.2%, 3.1%, and 1.9% in 2008, 2009, and 2011 seasons, respectively. While there was no enhancement for classification models based on sucrose for

the 2008 season when fusing sensor data, increase of 2.9% and 5.8% were found for the 2009 and 2011 seasons, respectively. It is worth comparing the classification results in this study to those obtained for whole potato tubers based on individual cultivars (Frito Lay 1879 or FL and Russet Norkotah or RN). In the earlier work, using an interactance system (446 to 1125 nm) and by applying the IPLS for wavelength selection, the classification accuracy values were as high as 100% and 86% for glucose and sucrose, respectively (Rady and Guyer 2015a). In another study, NIR reflectance was applied (900–1700 nm) on both cultivars and the optimal classification accuracy values were 100% for glucose and 79% for sucrose (Rady and Guyer 2015b). The earlier work had higher model classification accuracy, primarily as the models were developed on a single cultivar with less inherent variability in the tubers. The current study, however, illustrates an advantage over previous work as it has developed models that are not cultivar-specific. It is important to understand why the hyperspectral imaging performed better in general than spectroscopic systems in this study in both regression and classification tasks. Sugar distribution inside the potato tuber is generally not uniform and is a function of the cultivar, growing, and storage conditions (Pritchard and Scanlon 1997; Stark and Love 2003). Sucrose concentration

Table 6 Best classification models for glucose and sucrose of potato tubers using individual and fused data from, Vis/NIR interactance, Vis/NIR hyperspectral imaging, and NIR reflectance systems obtained from selected wavelengths for data combined from Frito Lay 1879 and Russet Norkotah cultivars based on testing the data over different seasons. Rows with values in boldface represent the best models for each season

Constituent	Season (training)	Season (testing)	Sensor	Preprocessing method (spectra, constituent)*	Classification technique	Classification accuracy (%)	
						Training set	Testing set
Glucose	2008	2009	VIS/NIR interactance	Non, Log	Knn	90.1	50.6
	2008	2011	VIS/NIR interactance	SNV, Log	Knn	92.6	60.6
	2009	2011	NIR reflectance	MSC, Log	Knn	78.2	54.3
	2011	2009	NIR reflectance	MC, Log	Knn	63.7	50.4
	2009	2011	VIS/NIR interactance + NIR reflectance	SNV, Log	Knn	62.5	60.6
	2011	2009	VIS/NIR interactance + NIR reflectance	1 st Derv, Log	LDA	62.1	59.8
Sucrose	2008	2009	VIS/NIR interactance	SNV, Log	Knn	74.8	68.4
	2008	2011	VIS/NIR interactance	SNV, Log	Knn	75.0	56.9
	2009	2011	NIR reflectance	MC, Log	Knn	70.3	67.3
	2011	2009	NIR reflectance	MC, Log	Knn	68.3	62.8
	2009	2011	VIS/NIR interactance + NIR reflectance	MC, Log	Knn	76.4	64.0
	2011	2009	VIS/NIR interactance + NIR reflectance	1 ST Derv, Log	Knn	71.3	65.4

* Non, no preprocessing; Log, logarithmic; SNV, standard normal variate; MSC, multiplicative signal correction; MC, mean centering; 1st Derv, smoothing using the first derivative

for two potato cultivars (Russet Burbank and Shepody) was higher towards the tuber center, whereas the glucose concentration was cultivar-dependent (Pritchard and Scanlon 1997). Thus, sampling techniques for assessing sugar content in the tubers as well as juicing for subsequent wet chemistry experiments play an important role in obtaining consistent results for regression or classification. Sampling was conducted in this study in a consistent way to eliminate any source of variation due to such factors. The HSI system provides greater spatial information of the chemical composition in the sample which yields a more accurate estimation of the desired constituent (Wu and Sun 2013), whereas spectroscopic systems are still based on a point measurement which does not cover a large sample area and cannot account for spatial variations in measured properties (Nicolai et al. 2007). Therefore, HSI systems can provide more robust and consistent information about the sugar concentration in the tubers and be used to develop better regression or classification results. It is also noted that interactance mode generally produced a better performance for regression and classification than reflectance mode. The possible reason is that the interactance mode is suitable for thin skin intact fruits as it eliminates the specular (i.e., surface) reflectance associated with the reflectance mode and it also provides more penetration to the sample tissue compared with the reflectance mode (Saranwong and Kawano 2007).

Regression and Classification Model Performance over Different Growing Seasons

The machine learning models developed in each growing season were tested with unseen data from the other seasons to determine their performance. The results from the regression models are in Table 5 and for the classification models in Table 6. It was observed that the 2008 interactance model when applied on the 2009 and 2011 data yielded r values of 90.3% and 92.1%, for glucose and 82.2% and 77.7% for sucrose, respectively, whereas applying the 2009 reflectance model on the 2011 season yielded an r value of 68.6% and the opposite case yielded 55%. The r values for sucrose were even lower (42.4 and 35.8% for 2009 and 2011, respectively). Fusing the interactance and reflectance data did not produce an improvement in the regression models. The possible reason for such a result was the relatively low performance obtained from the reflectance sensor.

In the case of classification, generally, the 2008 interactance system data as the training set had improved performance over the reflectance system data as shown in Table 6. Accuracy values for glucose models were 50.6% for 2009 and 60.6% for 2011, whereas the values were 68.4% and 56.9% in the case of sucrose. The reflectance data resulted in accuracy values of as high as 54.3% for glucose and 67.3% for sucrose. Data fused from the interactance and

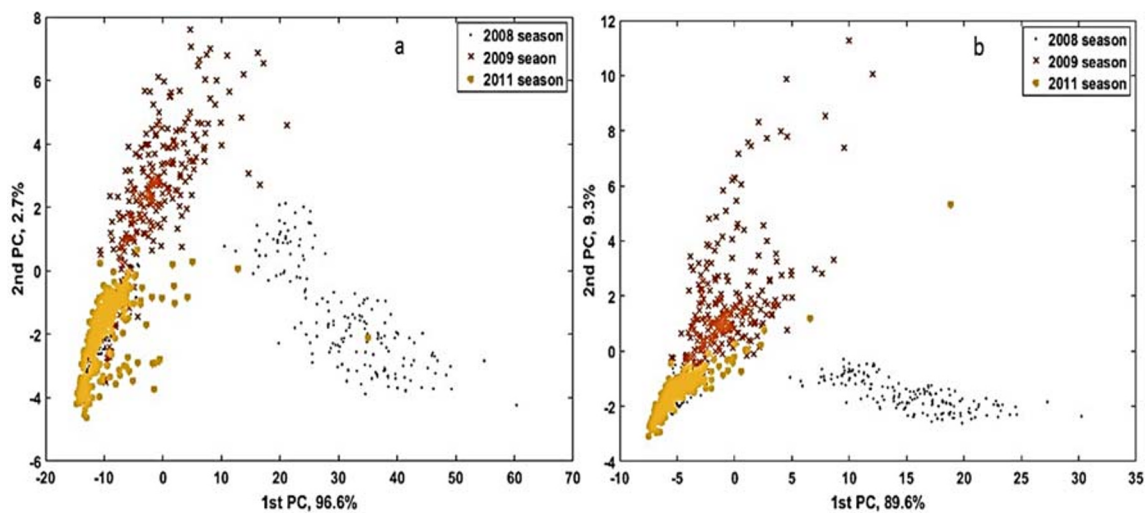


Fig. 5 PCA score plot for the interactance data obtained for potato tubers over the 2008, 2009, and/or 2011 seasons for **a** glucose and **b** sucrose

reflectance did not increase the classification accuracy and the highest accuracy was 62.5% for glucose and 76.4% for sucrose. Variability of the performance in regression or classification is mainly due to the different sugar ranges that were obtained in the three seasons as shown in Table 1. This was expected due to natural variation between the batches of tubers and different storage conditions utilized in the three seasons. The results of models over the three seasons can be improved by either including more samples, data from more seasons, or testing the models in work implemented in other cultivars. A score plot of the PCA obtained from the interactance data for the three seasons' data (i.e., 2008, 2009, 2011) is shown in Fig. 5a for glucose and Fig. 5b for sucrose. The data from the 2009 or 2011 seasons were less scattered than those for the 2008 season especially for glucose which shows the effect of various storage temperature or samples' sources.

Based on the results of the regression and classification operations performed in this study, the application of combining the data of multiple spectroscopic and hyperspectral imaging sensors to enhance the prediction of glucose and sucrose of whole potatoes is possible and beneficial for tracking quality parameters of tubers dedicated for frying, especially during cold storage.

Conclusions

This study has attempted to use data acquired from three spectroscopic systems, processed either individually or fused for building generic regression and classification models for glucose and sucrose content evaluation in potato tubers. These technologies could be used to monitor sugar content during storage and result in higher quality processed potato products. Data was collected over the three seasons, 2008, 2009, and

2011, and different machine learning algorithms were implemented for wavelength selection and model development. Classification models developed from the different sensors required fewer wavelengths than the regression models. In general, hyperspectral imaging presented superior classification and regression efficacy over interactance or reflectance systems due to the spatial data it records. Sensor fusion generally showed improved model accuracy performance over individual sensors for both regression and classification although the improvement was greater for the regression models. Considering the accuracy, it is advised to develop a device based on the hyperspectral sensor obtained in this study. However, the cost and time to acquire and process data from an HSI system are high and a limitation for industrial adoption. In this study, classification models were found to perform better than regression models except when using one season's model on data acquired from another season. It is worth stating that while regression models produce quantitative values of quality parameters, classification models are easier to develop (as they require less data) and could be more suitable for generating a rapid indication of potato condition.

The results obtained in this study have demonstrated how sensor measurements and machine learning can be used to evaluate the sugar content in potatoes and studied the effects of utilizing different data sets collected over different seasons. Although the model prediction results using data across different growing seasons were not as good as those developed for a single season, they could be improved by including data from more seasons and different cultivars. For industrial adoption of these techniques, it is important to develop models that can be applied across multiple seasons as it eliminates the cost of developing models every season and enables the models to be used immediately in a specific season and not until enough data is collected to develop a robust model.

Acknowledgments The authors acknowledge the USDA-ARS Postharvest Lab of Dr. Renfu Lu, the MSU Plant Pathology Lab of Dr. Dennis Fulbright, Sara Stadt at MSU Plant Pathology Lab, Rob Shafer, Walther Farms, Michigan, and the MSU AgBioResearch.

Funding This project was partially funded by the USDA-ARS-State Partnership Potato Program.

Compliance with Ethical Standards

Conflict of Interest Ahmed Rady declares that he has no conflict of interest. Daniel Guyer declares that he has no conflict of interest. Nicholas Watson declares that he has no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Amgren M, Hansen PW, Eriksen B, Larsen J, Larsen R (2011) Analysis of pregerminated barley using hyperspectral image analysis. *J Agric Food Chem* 59:11385–11394
- Ballabio D, Consonni V (2013) Classification tools in chemistry. Part 1: linear models. PLS-DA. *Anal Methods* 5:3790–3798
- Bohl, WH, Johnson, SB (2010) Commercial potato production in North America. The Potato Association of America handbook, second revision of American potato journal supplement 57, 38–48
- Camps C, Camps Z-N (2019) Optimized prediction of reducing sugars and dry matter of potato frying by FT-NIR spectroscopy on peeled tubers. *Molecules* 24:967
- Chen JY, Zhang H, Miao Y, Asakura M (2010) Nondestructive determination of sugar content in potato tubers using visible and near infrared spectroscopy. *Jpn J Food Eng* 11:59–64
- Chen JY, Zhang H, Miao Y, Matsunaga R (2005) NIR measurement of specific gravity of potato. *Food Sci Technol Res* 11:26–31
- Christy AA, Kvalheim OM (2007) Latent-variable analysis of multivariate data in infrared spectrometry. In: Ozaki Y, McClure WF, Christy AA (eds) *Near-infrared spectroscopy in food science and technology*. Wiley-Interscience, USA, pp 145–162
- Diezma B, Lleó L, Roger JM, Herrero-Langreo A, Lunadei L, Ruiz-Altisent M (2013) Examination of the quality of spinach leaves using hyperspectral imaging. *Postharvest Biol Technol* 85:8–17
- Dixit Y, Casado-Gavaldá MP, Cama-Moncunill R, Cama-Moncunill X, Markiewicz-Keszycka M, Cullen P, Sullivan C (2017) Developments and challenges in online NIR spectroscopy for meat processing. *Compr Rev Food Sci Food Saf* 16:1172–1187
- Duda RO, Hart PE, Stork DG (2012) *Pattern classification*. Wiley, NY
- Ekramirad N, Rady A, Adedeji AA, Alimardani R (2017) Application of Hyperspectral imaging and acoustic emission techniques for apple quality prediction. *Trans ASABE* 60:1391–1401
- Fairuz Omar A (2013) Spectroscopic profiling of soluble solids content and acidity of intact grape, lime, and star fruit. *Sens Rev* 33:238–245
- Giovenzana V, Beghi R, Civelli R, Guidetti R (2016) Application of near infrared spectroscopy and development of simplified optical devices for the fresh-cut fruit and vegetable sector. *NIR News* 27:4–6
- Griffiths PR, Dahm DJ (2007) Multilinear regression and principal component analysis. In: Burns DA, Ciurezak EW (eds) *Handbook of near-infrared analysis*. CRC press, Boca Raton, p 21
- Hall DL, Llinas J (1997) An introduction to multisensor data fusion. *Proc IEEE* 85:6–23
- Hartmann R, Büning-Pfaue H (1998) NIR determination of potato constituents. *Potato Res* 41:327–334
- Heise HM, Winzenm R (2001) *Fundamental chemometric methods*. In: Siesler HW, Ozaki Y, Kawata S, Heise HM (eds) *Near-infrared spectroscopy: principles, instruments, applications*. Wiley-Vch Verlag GmbH., Weinheim, pp 125–162
- Helgerud T, Segtnan VH, Wold JP, Ballance S, Knutsen SH, Rukke EO, Afseth NK (2012) Near-infrared spectroscopy for rapid estimation of dry matter content in whole unpeeled potato tubers. *J Food Res* 1:55
- Huang M, Wan X, Zhang M, Zhu Q (2013) Detection of insect-damaged vegetable soybeans using hyperspectral transmittance image. *J Food Eng* 116:45–49
- Ignat T, Alchanatis V, Schmilovitch Z (2014) Maturity prediction of intact bell peppers by sensor fusion. *Comput Electron Agric* 104:9–17
- Kemps B, Leon L, Best S, De Baerdemaeker J, De Ketelaere B (2010) Assessment of the quality parameters in grapes using VIS/NIR spectroscopy. *Biosyst Eng* 105:507–513
- Li G, Wang R, Quampah AJ, Rong Z, Shi C, Wu J (2011) Calibration and prediction of amino acids in stevia leaf powder using near infrared reflectance spectroscopy. *J Agric Food Chem* 59:13065–13071
- Li ML, Ekramirad N, Rady A, Adedeji A (2018) Application of acoustic emission and machine learning to detect codling moth infested apples. *Transactions of the ASABE* 61(3):1157–1164. <https://doi.org/10.13031/trans.12548>
- Liu Y, Sun X, Ouyang A (2010) Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN. *LWT-Food Sci Technol* 43:602–607
- Lohumi S, Lee S, Lee H, Cho B-K (2015) A review of vibrational spectroscopic techniques for the detection of food authenticity and adulteration. *Trends Food Sci Technol* 46:85–98
- Manso JY (2008) *Sensor fusion of IR, NIR, and Raman spectroscopic data for polymorph quantitation of an agrochemical compound*. Doctoral dissertation, University of Delaware.
- Mao KZ (2004) Orthogonal forward selection and backward elimination algorithms for feature subset selection. *IEEE Transact Syst Man Cybernet Part B (Cybernetics)* 34:629–634
- Mark H (2001) Data analysis: multilinear regression and principal component analysis. In: Burns DA, Ciurezak EW (eds) *Handbook of near-infrared analysis*. Marcel Dekker Inc, New York, pp 129–184
- Mendoza F, Lu R, Ariana D, Cen H, Bailey B (2011) Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content. *Postharvest Biol Technol* 62:149–160
- Mendoza F, Lu R, Cen H (2012) Comparison and fusion of four nondestructive sensors for predicting apple fruit firmness and soluble solids content. *Postharvest Biol Technol* 73:89–98
- Mottram DS, Wedzicha BL, Dodson AT (2002) Food chemistry: acrylamide is formed in the Maillard reaction. *Nature* 419:448–449
- Nicolai BM, Beullens K, Bobelyn E, Peirs A, Saeys W, Theron KI, Lammertyn J (2007) Nondestructive measurement of fruit and

- vegetable quality by means of NIR spectroscopy: a review. *Postharvest Biol Technol* 46:99–118
- NPC (2019) 2019 Potato Statistical Yearbook. National Potato Council Capital Hilton, Washington, D.C.
- Peng Y, Lu R (2008) Analysis of spatially resolved hyperspectral scattering images for assessing apple fruit firmness and soluble solids content. *Postharvest Biol Technol* 48:52–62
- Pritchard M, Scanlon M (1997) Mapping dry matter and sugars in potato tubers for prediction of whole tuber process quality. *Can J Plant Sci* 77:461–467
- Qu JH, Liu D, Cheng JH, Sun DW, Ma J, Pu H, Zeng XA (2015) Applications of near-infrared spectroscopy in food safety evaluation and control: a review of recent research advances. *Crit Rev Food Sci Nutr* 55:1939–1954
- Rady A, Adedeji A (2018) Assessing different processed meats for adulterants using visible-near-infrared spectroscopy. *Meat Sci* 136:59–67
- Rady A, Fischer J, Reeves S, Logan B, James Watson N (2020) The effect of light intensity, sensor height, and spectral pre-processing methods when using NIR spectroscopy to identify different allergen-containing powdered foods. *Sensors* 20:230
- Rady A, Guyer D (2015a) Utilization of visible/near-infrared spectroscopic and wavelength selection methods in sugar prediction and potatoes classification. *J Food Measurem Characterizat* 9:20–34
- Rady A, Guyer D, Lu R (2015) Evaluation of sugar content of potatoes using hyperspectral imaging. *Food Bioprocess Technol* 8:995–1010
- Rady AM, Guyer DE (2015b) Evaluation of sugar content in potatoes using NIR reflectance and wavelength selection techniques. *Postharvest Biol Technol* 103:17–26
- Rady AM, Guyer DE (2015c) Rapid and/or nondestructive quality evaluation methods for potatoes: a review. *Comput Electron Agric* 117:31–48
- Rady AM, Guyer DE, Kirk W, Donis-González IR (2014) The potential use of visible/near infrared spectroscopy and hyperspectral imaging to predict processing-related constituents of potatoes. *J Food Eng* 135:11–25
- Rama M, Narasimham P (2003) Potatoes and related crops fruits of the solanaceae. In: Caballero B (ed) *Encyclopedia of food science and nutrition*. Academic Press, USA, pp 4658–4680
- Sahni NS, Isaksson T, Næs T (2004) In-line near infrared spectroscopy for use in product and process monitoring in the food industry. *J Near Infrared Spectrosc* 12:77–83
- Saranwong S, Kawano S (2007) Fruits and vegetables. In: Ozaki Y, McClure WF, Christy AA (eds) *Near-infrared spectroscopy in food science and technology*. Wiley-Interscience, USA, pp. 219–245
- Serranti S, Cesare D, Marini F, Bonifazi G (2013) Classification of oat and groat kernels using NIR hyperspectral imaging. *Talanta* 103:276–284
- Shao Y, He Y, Mao J (2006) Nondestructive measurement of the internal quality of bayberry juice using NIR spectroscopy. *Food Eng* 79(3):1015–1019. <https://doi.org/10.1016/j.jfoodeng.2006.04.006>
- Singh C, Paliwal J, Jayas D, White N (2006) Near-infrared spectroscopy: applications in the grain industry. The CSBE/SCGAB Annual Conference, Edmonton, Alberta, Canada
- Stadler R, Blank B, Varga N, Robert F, Hau J, Guy P, Robert M, Riediker S (2002) Acrylamide from Maillard reaction products. *Nature* 419:449–450
- Stark JC, Love SL (2003) Tuber quality. *Potato production systems*. University of Idaho, Aberdeen, pp 329–343
- Storey R, Davies H (1992) Tuber quality. In: Harris P (ed) *The potato crop*. Springer, Dordrecht, pp 507–569. <https://doi.org/10.1007/978-94-011-2340-2>
- Subedi PP, Walsh KB (2009) Assessment of potato dry matter concentration using short-wave near-infrared spectroscopy. *Potato Res* 52:67–77
- Ulloa PA, Guerra R, Cavaco AM, Da Costa AMR, Figueira AC, Brigas AF (2013) Determination of the botanical origin of honey by sensor fusion of impedance e-tongue and optical spectroscopy. *Comput Electron Agric* 94:1–11
- Varmuza K, Filzmoser P (2016) *Introduction to multivariate statistical analysis in chemometrics*. CRC Press, Boca Raton
- Workman J Jr, Weyer L (2012) *Practical guide and spectral atlas for interpretive near-infrared spectroscopy*. CRC press, Boca Raton, pp 89–92. <https://doi.org/10.1201/b11894>
- Wu D, Sun D-W (2013) *Hyperspectral imaging technology: a nondestructive tool for food quality and safety evaluation and inspection*, Advances in food process engineering research and applications. Springer, Boston pp 581–606
- Yaptenko K, Suzuki T, Kawakami S, Sato H, Takano K, Kozima T (2000) Nondestructive determination of sugar content in ‘Danshaku’ potato (*Solanum tuberosum* L.) by near infrared spectroscopy. *J Agric Sci Tokyo University of Agriculture* 44(4):284–294
- Zhang S, Zhang H, Zhao Y, Zhao H (2012) Comparison of modeling methods of fresh jujube soluble solids measurement by NIR spectroscopy. *Nongye Jixie Xuebao= Transactions of the Chinese Society for Agricultural. Machinery* 43:108–112

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.