

Algorithms for detecting cherry pits on the basis of transmittance mode hyperspectral data

Anna Siedliska*, Monika Zubik, Piotr Baranowski, and Wojciech Mazurek

Institute of Agrophysics, Polish Academy of Sciences, Doświadczalna 4, 20-290 Lublin, Poland

Received March 27, 2017; accepted September 5, 2017

Abstract. The suitability of the hyperspectral transmittance imaging technique was assessed in terms of detecting the internal intrusions (pits and their fragments) in cherries. Herein, hyperspectral transmission images were acquired in the visible and near-infrared range (450-1000 nm) from pitted and intact cherries of three popular cultivars: ‘Lutówka’, ‘Pandy 103’, and ‘Groniasta’, differing by soluble solid content. The hyperspectral transmittance data of fresh cherries were used to determine the influence of differing soluble solid content in fruit tissues on pit detection effectiveness. Models for predicting the soluble solid content of cherries were also developed. The principal component analysis and the second derivative pre-treatment of the hyperspectral data were used to construct the supervised classification models. In this study, five classifiers were tested for pit detection. From all the classifiers studied, the best prediction accuracies for the whole pit or pit fragment detection were obtained via the back-propagation neural networks model (87.6% of correctly classified instances for the training/test set and 81.4% for the validation set). The accuracy of distinguishing between drilled and intact cherries was close to 96%. These results showed that the hyperspectral transmittance imaging technique is feasible and useful for the non-destructive detection of pits in cherries.

Keywords: cherry, hyperspectral transmittance, pit detection, soluble solid content

INTRODUCTION

Tart cherries (*Prunus cerasus*) are greatly appreciated by consumers for their taste and nutritional properties. The European Union is a major world producer of sour cherries, with a total of 655,000 MT of production for the 2016 season (FAOSTAT, 2016). Most of the produced cherries are processed and sold frozen or canned, with a majority being pitted (USDA, 2016). The fruit quality classification is

currently based on the evaluation of the surface properties of the fruits, including surface colour, shape and defects. However, in cherries processing, there are many internal features, such as pit presence, decay and insect damage, which can have impact on both the quality and safety of these products (Donis-González *et al.*, 2015; Xing *et al.*, 2008). Effectively sorting out these undesirable features during handling and processing is an underlying issue determining the quality of the final product. Hence, new methods, including computed tomography (CT), x-ray imaging or near infrared spectroscopy, are being tested for detecting external and internal inhomogeneities within fruits tissue (Kawano, 2016; Nicolaï *et al.*, 2014).

A variety of techniques have been reported for pit and pit fragments detection in stone fruits (Allen *et al.*, 1996), and in other commodities with seeds, such as olives (Zion *et al.*, 1997) and dried plums (Haff *et al.*, 2005). Moreover, considerable development was achieved in attempts to detect pits and their fragments in cherries (Haff *et al.*, 2013). Timm *et al.* (1991) studied different methods, including microwave transmission, ultrasound reflection, light-beam interruption, light-beam transmission and machine vision for pit detection. They demonstrated that light transmission coupled with machine vision and image analysis was able to detect 95% of the cherries with pits. Even better results (97% accuracy in classifying both pitted and unpitted cherries) were obtained using nuclear magnetic resonance (NMR) by Zion *et al.* (1994). The NMR has been also used to identify pits in olives (Zion *et al.*, 1997). The accuracy of this method was 97%. However, it is costly and unlikely to be adopted by the industry. Haff *et al.* (2013) used

*Corresponding author e-mail: a.siedliska@ipan.lublin.pl

a one dimensional x-ray detector array for detecting pits in cherries. This method was found to be 97.3% accurate in classifying pitted cherries, and 94% in classifying unpitted fruits. Unfortunately, the x-ray imaging method is also expensive and needs a robust electrical energy supply (Haff and Toyofuku, 2008). In recent years, CT has also been tested for detection of pit and pit fragments in fruits. For this purpose, agarose phantoms were developed by Donis-González *et al.* (2015). The authors have demonstrated that the technique in question allows the distinguishing of pit and pit fragments with a high accuracy rate ($R=0.99$). Still, because of the economic cost and lack of speed, CT imaging has proven to be non-practical for being implemented as a real time food inspection technology.

In agriculture, hyperspectral imaging (HSI) systems using appropriate illumination patterns, such as reflectance, transmittance and interactance (Pan *et al.*, 2017; Szuvandzsiev *et al.*, 2014), have been successfully demonstrated to be accurate enough for the routine inspection of internal quality parameters in food products (Lu *et al.*, 2017). Recently, this technology has also been commonly employed to assess the soluble solid content (SSC), titrable acidity (TA), pH, and firmness in hami melons (Sun *et al.*, 2016), and persimmon (Munera *et al.*, 2017). Moreover, several studies have demonstrated the feasibility of HSI for detecting defects in fruits and vegetables (Baranowski *et al.*, 2013; Pu *et al.*, 2015). Song *et al.* (2016) conducted a study to detect black heart in white radish. Based on 4 selected wavelengths and using Fisher's linear discriminant analysis (FLDA), 98.4% discrimination accuracy of defect radishes was obtained. In addition, cold injury in peaches was successfully detected using HSI and an artificial neural network (ANN) model (Sun *et al.*, 2017). The overall classification accuracy of chill damage was 95.8%.

Therefore, the objective of this research was to examine the applicability of hyperspectral imaging in the visible and near-infrared (VNIR) wavelength range for detecting pits in fresh cherries of three selected cultivars. The specific aims of this study were to: (1) evaluate the applicability of hyperspectral transmittance imaging over the spectral range of 400-1000 nm of fresh tart cherries, so as to detect pits and pit fragments; (2) identify the optimal wavelengths for pit and pit fragments detection; (3) develop classification models based on selected wavelengths, and compare the classification performance among different classifiers; (4) investigate the feasibility of using VNIR hyperspectral transmission data to predict SSC contents of different varieties of cherry.

MATERIALS AND METHODS

For the research purpose, 540 cherry fruits belonging to three cherry (*Cerasus vulgaris* Mill.) cultivars: 'Łutówka', 'Pandy 103' and 'Groniasta' were purchased from a local fruit and vegetable cold storage plant (Fructosad, in

Ratoszyn, Poland). The cultivars were selected based on their difference in physical properties, especially SSC (Wojdyło *et al.*, 2014). The samples had similar sizes and shapes, and were free of external defects (no bruising, cracking or signs of diseases) based on visual inspection. No significant differences in colour were observed within each cultivar. It was expected that differences in cherry size could strongly influence pit detection (Qin and Lu, 2005). Therefore, cherries with a large range of sizes (diameter ranging from 15-22 mm) were used in each variant of the experiment, representing 99% of all the available cherries. Cherries from each cultivar were divided into three groups, each containing 60 cherries: cherries with whole pit, cherries with pit fragments, and cherries without a pit. The pit removal was done with the use of a commercial hand-held cherry pitter with a plunger diameter of 3 mm. To create a variant with pit fragments, the pits removed from the cherries were crushed with a hammer. The areas of pit fragments and their highest lengths were determined using an average of three measurements of their sizes with a calliper. Such measurements were calibrated for selected pit fragments on the base of 3D scanning with microtomograph GE Nanotom 180 (GE Sensing and Inspection Technologies GmbH, Wunstorf, Germany). For the 3D analysis of the one side area, VG Studio MAX 2.1 software was used. Only fragments with the one side (outer surface) area between 4 mm² and 8 mm² were selected. Another criterion for pit fragment selection was the distance between two most distant points on the pit fragment surface being not higher than 3 mm. The smallest outer surface area of the pit fragment of 4 mm² was chosen in the study because such fragments were readily detectable by the hyperspectral system. In one drilled cherry fruit, one pit fragment was put inside. During the experiment, fresh samples were stored at 5°C and removed from cold storage about two hours before hyperspectral imaging to allow them to reach room temperature.

The hyperspectral transmittance imaging system used in this study consisted of a linear hyperspectral scanner composed of a VNIR camera with an ImSpector V10E imaging spectrograph for the spectral range of 400-1000 nm. The device was manufactured by SPECIM, Finland. A transmittance illumination unit consisted of two linear halogen lamps of 100 W each. The lamps were housed in an aluminium enclosure installed below the table and positioned on the belt. The fruit samples were placed in two rows on a positioning table having six holes, each 15 mm in diameter. The light source illuminated the cherry from below in the vertical direction towards the detector.

Hyperspectral transmission images were obtained for all cherries belonging to nine variants. The exposure time was set at 16 ms for obtaining good quality images without saturation. The spatial dimension of all the images acquired in this work was 696 x 519 pixels; a total of 519 images were collected in the spectral range of 450-1000 nm after

performing 2 x 2 binning operations in the spatial and spectral directions. The measurements were performed in a dark room to prevent the influence of external illumination.

In order to assess the influence of the internal fruit quality on transmission spectra, the SSC was determined using traditional destructive tests. After the image acquisition, each cherry was cut in half. Juice from both cherry halves was extracted, and the SSC (%) was measured using a PAL-BX/RI pocket refractometer produced by Atago Co., Ltd., Tokyo, Japan.

Hyperspectral images were analysed to extract information about the spectral properties of both pitted and unpitted fruits, for the purpose of optimizing pit and pit fragments detection, selecting the effective wavelengths and applying the adequate classification. After image acquisition, the areas containing regions of interest (ROI), *i.e.* regions including the area of whole cherry fruit with a pit, pit fragment, or without a pit, were identified from the hyperspectral cube. To distinguish the ROI areas, a script was written using ImageJ software. The segmentation procedure was a modification of the procedure proposed earlier by Baranowski *et al.* (2013). In the first step, one image was selected to create a binary mask of the fruit (herein, the 762 nm wavelength image was used due to the high contrast between fruit surface and background). This binary mask of the fruit surface was subsequently applied to all images in the hyperspectral cube to eliminate the background. Each selected ROI consisted of about 15 000 pixels. The average transmittance spectrum from the ROI of pitted and intact cherries was computed by averaging the spectral value of all pixels in the ROI for each segmented cherry image. In this way, one average spectrum was obtained for each cherry sample. The mean transmittance spectra calculated from the ROI of fruits with pits were compared with the same fruit after the pitting procedure (the fruit that does not contain a pit) and with the variant containing pit fragments.

The pre-treatment of spectral characteristics was completed with the use of the Unscrambler X ver. 10.1, CAMO Software, manufactured in Oslo, Norway. Prior to modelling, the pre-treatment of the mean spectral data consisted of the application of an automatic baseline correction and a second derivative calculated using the Savitzky-Golay method. The automatic baseline correction was chosen to remove the baseline effects from the spectra which occurred during spectral collection. The derivative processing effectively reduced the image artefacts caused by non-uniform illumination and also eliminated the background signal. The effectiveness of this pre-treatment method was confirmed by Siedliska *et al.* (2014) and Baranowski *et al.* (2015). In order to reduce redundancy and co-linearity in the hyperspectral data collected, as well as to improve the efficiency of pit detection and to meet the inspection speed required by industry, wavelength selection was a critical and important step. In this study, two different methods were used to select key wavelengths. In the first method, the appropriate

data were selected as attributes from peaks in the second derivative spectral characteristics curves. Peaks occurring in the second derivative curve show the curvature of the original spectra. The most characteristic feature of the second derivative is a negative band with a minimum at the same wavelength as the maximum in the zero-order band. There are also two additional positive bands on either side of the negative peak (Becker *et al.*, 2005). Based on this method, 27 wavelengths were selected as attributes from the peaks in the spectral characteristics second derivative. These were then used to construct the supervised classification models. The second method for reducing the number of data features in the spectral data was the principal component analysis (PCA) used for the smoothed spectra after second derivative pre-treatment. The method is based on the fact that neighbouring bands of hyperspectral images are highly correlated and often convey the same information about the object. In this method, the original data are transformed to remove correlation among the bands. This process is based on identification of the optimum linear combination of the original bands accounting for the variation of pixel values in any particular image. This multivariate statistical technique used for the whole spectral data cube allowed a decrease in the large number of potential wavelengths, to several non-correlated principal components (PCs). A similar methodology, used as the pre-processing method before supervised classification, was tested by other authors (Liaghat *et al.*, 2014). It is recommended that the number of PCs (spectral features) should be selected such that the PCs represented >90% variability within the original data. In this study, the first PCA components were used as independent variables in the classification models of pit detection. They covered 97.2% of all variability within the data. Data analysis and classification were then performed on the transmittance hyperspectral data of the fruit samples. In this study, for the supervised classification methods, cherry samples were randomly selected into training and test sets consisting respectively of 90 and 10% of all the test population. The training data set was used to build the classification model, while the test data set was used to test its capability for classifying new samples. The learning and testing procedure was repeated ten times. For each new run, the instances selected as training and test sets were changed (cross-validation method). The stratified 10-fold cross-validation method has been proven as the standard evaluation technique in the events where only limited data is available. This technique is also regarded as being the most rigorous (Witten and Frank, 2005). In this study, a supervised learning approach was employed to verify the classification algorithm. All the classification algorithms were implemented using comprehensive software called the Waikato Environment for Knowledge Analysis, or Weka. This software contains tools for data pre- and post-processing, and for evaluating the results of learning schemes. Pre-processing of the hyperspectral data

Table 1. Chosen features of the classifiers used in the study

| Name of classifier's library | Description of algorithm | Acronym | Chosen parameters of classifier |
|------------------------------|--|---------|---|
| Naive Bayes | Naive Bayes classifier which uses estimator classes. Numeric estimator precision values are chosen based on analysis of the training data. | NB | Use Supervised Discretization: true, Debug: false, Use Kernel Estimator: false |
| LibSVM | A wrapper class for the libsvm tools. Allows users to experiment with One-class SVM, Regressing SVM, and nu-SVM supported by LibSVM tool. | SVM | SVM Type: nu-SVC, Kernel Type: linear; Ny: 0.5, Normalize: true, Probability Estimates: true |
| Logistic | Builds linear logistic regression models | LLR | Debug: false, MaxIts: -1, Ridge: 1.0E-8 |
| Multilayer Perceptron | Uses backpropagation neural networks to classify instances. | BNN | AutoBuild: true, Learning rate: 0.3, Momentum: 0.2, Training time: 500 Hidden layers = (attribs + classes) / 2 |
| Random Forests | Classifier for constructing a forest of random trees. | RF | Debug: false, MaxDepth: 0, Num of Features : 0, Num of Trees: 10, Seed : 1 |

consisted of choosing, from the whole spectral range, the range where spectral characteristics of the signal were sufficiently strong. Due to the fact that transmittance for cherry samples at 600 nm or lower proved relatively low when compared to those between 600-1 000 nm, only the wavelength range from 600-1 000 nm was used for the analysis. Of the different classifiers available in Weka, the five with the best prediction accuracies were chosen for comparison. These classifiers are presented in Table 1, along with a general description and the actual parameters determined by this study.

The prediction models of SSC for the three cherry cultivars studied were created using the Weka Knowledge Flow Interface. This software was used for developing the backpropagation neural networks (BNN) and principal component analysis-backpropagation neural networks (PCA-BNN) models. The BNN model was created using wavelengths obtained as attributes from the peaks in the second derivative of the spectral characteristics of pitted cherries belonging to the three studied cultivars. For the PCA-BNN model, the PCA was used to extract information from the entire spectral range. All five principal components were used as neurons of the network input layers for building BNN. The quality of the models was evaluated using the coefficient of determination (R^2), the root mean square error of correlation (RMSEC), and the root mean square error of prediction (RMSEP). It is generally recom-

mended that a good model be characterized by high R^2 , low RMSEC and RMSEP, and also a small difference between RMSEC and RMSEP. In some publications, the standard error of prediction (SEP) is reported instead of the RMSEP. The difference between RMSEP and SEP was explained by Golic and Walsh (2006).

RESULTS AND DISCUSSION

The hyperspectral scanning of the cherries with transmittance mode enabled visualization of the presence of a pit within the tested fruits. However, the accuracy of pit detection depended on the size of the pit/pit fragments and the wavelength analyzed. The pit regions were reflected in the transmittance images as areas of lower intensity of transmitted light. To better visualize the presence of pit or lack of pit within cherries, a multiband composition images were created. An example of a representative three-band (650, 736, and 815 nm) hyperspectral transmittance composition of intact and pitted cherries of three studied cultivars is presented in Fig. 1. It can be noted in these images that, when cherries contain a pit or pit fragments, a darker area in the middle of the transmittance image of the fruit occurs, with an intensity differing for various cultivars and particular variants of the experiment (whole pit, pit fragment). The occurrence of the darker regions in the images of cherries with pit/pit fragments is the result of the higher optical density of a pit, as compared to fruit

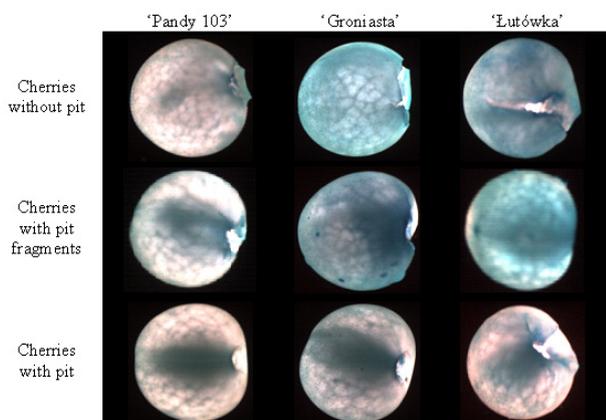


Fig. 1. Typical hyperspectral transmittance RGB images of intact and pitted cherries of the three cultivars studied (R: 650 nm, G: 736 nm, and B: 815 nm).

flesh. The size of this shadow may be influenced by the size of the pit and also by the orientation of the pit to the camera (Xing *et al.*, 2008). Moreover, the size of the fruit also has impact on the quality of transmittance images.

Although the three-band image composition of raw spectral data enabled detection of pitted cherries in many cases (Fig. 1), the classical segmentation methods applied to such images was seen to be insufficient for distinguishing cherries with and without a pit, with a satisfactory precision. Therefore, multiple variables analysis was undertaken. To present the characteristic peaks and troughs more clearly, the moving average smoothing procedure and the Savitzky-Golay second order derivative were applied to the raw transmittance data of the all studied variables. This procedure enhanced the features of the spectra by separating out peaks of overlapping bands and induced noise filtering of the spectral data. Several interesting spectral features are

apparent in the derivative spectra that were unnoticeable in the original spectra. Interestingly, the highest changes in curvature occurred between 680-780 nm, while the subtle changes proved consistent over the entire spectra. The main issue in the study was to distinguish cherries with pit and pit fragments from cherries without pits. In Fig. 2, the differences between the mean spectra of pitted cherries and cherries with pit/pit fragments, treated with second derivative, are presented for the three cultivars studied. In these plots, considerable differences of transmittance spectra are observed, especially at peaks and troughs. Based on the second derivative transformation of original spectra, the 27 bands that were situated at maxima or minima were selected and marked as dotted vertical lines in Fig. 2. These wavelengths were used in the supervised classification to distinguish intact and pitted cherries. All the selected wavelengths belong to the VNIR region of spectrum. This relates to the vibration and combination overtones of the C-H, O-H, and N-H bonds in organic molecules. Peaks in the range 660-700 nm were recognized as being the spectra bands of the anthocyanin pigments, representing the red colour of cherry fruits. The absorption peak at approximately 740 nm could be attributed to the third overtone of the O-H functional group in water, and to the fourth overtone of the C-H bond. The absorption peaks at approximately 830 nm and 918 nm were likely attributable to the third overtone of the C-H stretching vibration in sugar (Baiano *et al.*, 2012). Williams and Norris (2001) found sugar absorption bands at approximately 888 nm and 913 nm, and the peak that occurred at about 980 nm was assigned to the effect of water absorption, and corresponded to the second overtones of O-H stretching. Since the wavelengths selected for the classification procedure employed are highly correlated with SSC in fruits, it can be assumed that SSC can affect the precision of pit/pit fragments detection.

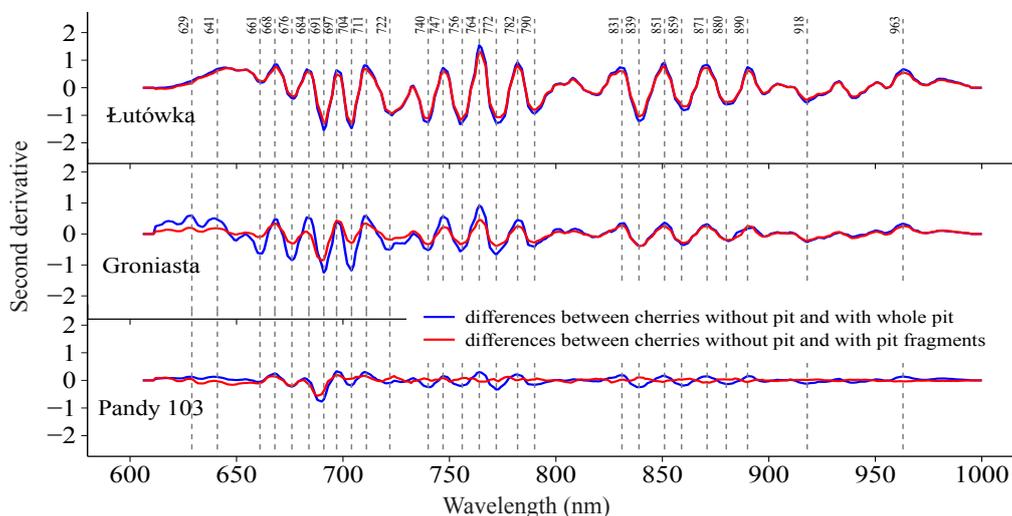


Fig. 2. Second derivative differential spectra of the three cultivars studied. The dotted vertical line represents the 27 selected spectral bands situated at maxima or minima.

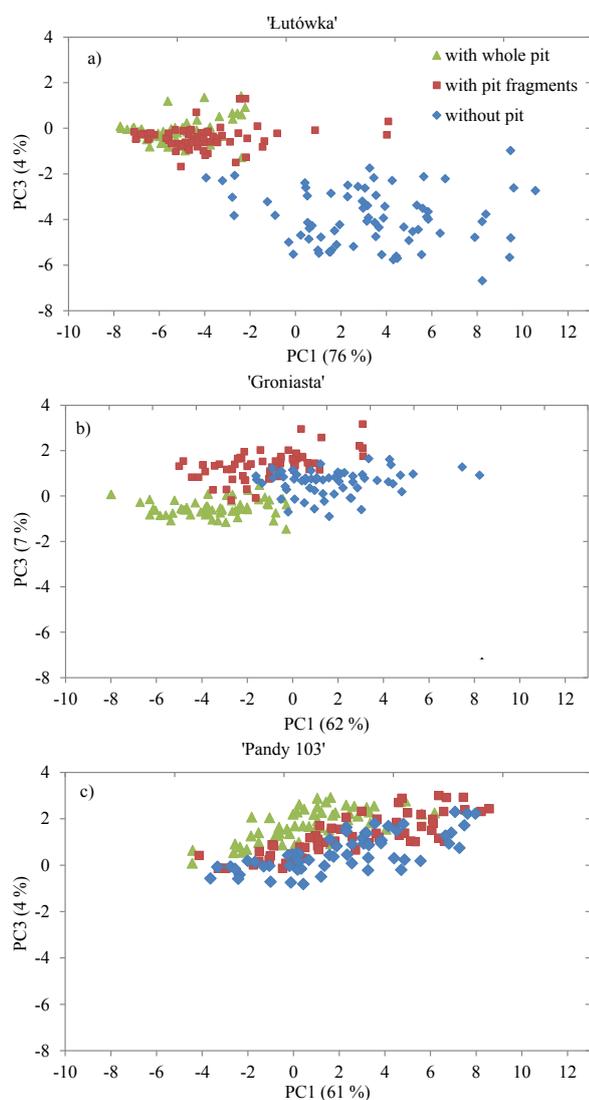


Fig. 3. Score plots of the first and third principal components of transmittance spectra of intact and pitted cherries of the three cultivars studied. Spectra were pre-processed by use of second derivative transformation using a Savitzky-Golay algorithm.

In the second method, dimensionality reduction was performed by transforming wavelength variables into principal components. Subsequently, PCA was applied to the smoothed spectral data after the second derivative pre-treatment. Figure 3 presents the score plot of PCA (PC1 and PC3) as applied to the dataset including pitted and intact cherry fruits. The first and third principal components showed the largest variation in the data. The PCA score plot of ‘Groniasta’ cultivar indicates that Groniasta samples can be clustered into three classes (cherries without pit, with pit fragments and with whole pit) with the presence of some overlap (Fig. 3a). For the ‘Łutówka’ cultivar (Fig. 3b), the score plot displayed a clear separation of cherries without pit from other groups. The differentiation between classes for ‘Pandy 103’ cultivar was, however, not so clear (Fig. 3c). Still, in this case, two classes: cherries without pit and cherries with whole pit could be separated.

The final results of comparing classification models used for distinguishing between pitted cherries, intact cherries, and cherries with pit fragments, as based on 27 wavelengths, are presented in Table 2. For all the classification algorithms studied, BNN had the best prediction accuracies (87% of all correctly classified instances for both training set and validation set). However, most of the tested models had prediction accuracy close to 75% of all correctly classified instances for the training set. From all the models studied, the worst classifier was the Naive Bayes (NB) model - with 58% of all correctly classified instances. Thus, BNN and RF models could be chosen as suitable classifiers for distinguishing between particular cultivars of cherries.

The second classification experiment was performed to distinguish between pitted cherries, intact cherries, and cherries with pit fragments, as based on second derivative data, taking into account five first PCA components. The classification performance results are summarized in Table 3. From all the classifiers presented in Table 1, the best overall classifier was obtained for the Random Forests (RF) model (85% of all correctly classified instances for the training/test set and 91% for the validation set). Similar

Table 2. Results of classification models for fresh pitted, intact cherries and cherries with pit fragments. The models were created based on the wavelengths selected as attributes from the peaks in the second derivative of spectral characteristics

| Classification model* | Training/ Test set | | | Validation set | | |
|-----------------------|------------------------------------|-----------------|------------------------|------------------------------------|-----------------|------------------------|
| | Correctly classified instances (%) | Kappa statistic | Root mean square error | Correctly classified instances (%) | Kappa statistic | Root mean square error |
| BNN | 87.65 | 0.81 | 0.25 | 81.48 | 0.72 | 0.29 |
| RF | 85.88 | 0.79 | 0.26 | 90.74 | 0.86 | 0.24 |
| SVM | 82.74 | 0.74 | 0.34 | 88.89 | 0.83 | 0.27 |
| LLR | 79.22 | 0.69 | 0.32 | 85.19 | 0.78 | 0.29 |
| NB | 68.63 | 0.53 | 0.38 | 57.41 | 0.36 | 0.45 |

*Classification model for distinguish between pitted and unpitted cherries.

Table 3. Results of classification models for fresh pitted cherries, intact cherries and cherries with pit fragments. The models were created based on second derivative data, taking into account the 5 first PCA components

| Classification model* | Training/ Test set | | | Validation set | | |
|-----------------------|------------------------------------|-----------------|------------------------|------------------------------------|-----------------|------------------------|
| | Correctly classified instances (%) | Kappa statistic | Root mean square error | Correctly classified instances (%) | Kappa statistic | Root mean square error |
| BNN | 87.25 | 0.81 | 0.26 | 87.04 | 0.81 | 0.26 |
| RF | 75.49 | 0.63 | 0.34 | 70.37 | 0.55 | 0.30 |
| SVM | 72.94 | 0.59 | 0.36 | 72.22 | 0.58 | 0.35 |
| LLR | 74.31 | 0.61 | 0.34 | 75.92 | 0.64 | 0.32 |
| NB | 58.23 | 0.37 | 0.50 | 53.70 | 0.31 | 0.54 |

*Classification model for distinguish between pitted and unpitted cherries.

Table 4. Confusion matrices obtained with 10-fold cross-validation for the BNN-PCA model (a) and for the BNN model (b) based on 27 selected wavelengths, for the studied variants as an independent variable. Each cultivar has a specific colour representation and the diagonal cells present the correct classification

| | With pit | | | With pit fragments | | | Without pit | | |
|--------------------|----------|-----------|-------|--------------------|-----------|-------|-------------|-----------|-------|
| | Łutówka | Groniasta | Panda | Łutówka | Groniasta | Panda | Łutówka | Groniasta | Panda |
| BNN-PCA model | | | | | | | | | |
| With pit | 58 | 59 | 45 | 2 | 1 | 15 | 0 | 0 | 0 |
| | 96.7% | 98.3% | 75% | 3.3% | 1.7% | 25% | 0% | 0% | 0% |
| With pit fragments | 1 | 0 | 7 | 57 | 59 | 53 | 1 | 1 | 0 |
| | 1.6% | 0% | 11.7% | 95% | 98.3% | 88.3% | 1.6% | 1.6% | 0% |
| Without pit | 0 | 0 | 1 | 0 | 2 | 1 | 60 | 58 | 58 |
| | 0% | 0% | 1.6% | 0% | 12.5% | 1.6% | 100% | 96% | 96.7% |
| BNN model | | | | | | | | | |
| With pit | 59 | 59 | 44 | 1 | 1 | 16 | 0 | 0 | 0 |
| | 98.3% | 98.3% | 73.3% | 1.7% | 1.7% | 26.7% | 0% | 0% | 0% |
| With pit fragments | 2 | 0 | 7 | 58 | 60 | 53 | 0 | 0 | 0 |
| | 3.3% | 0% | 11.7% | 96.7% | 100% | 88.3% | 0% | 0% | 0% |
| Without pit | 1 | 0 | 1 | 0 | 0 | 0 | 59 | 60 | 59 |
| | 1.7% | 0% | 1.7% | 0% | 0% | 0% | 98.3% | 100% | 98.3% |

accuracies of 88 and 83% of all correctly classified instances for the training/test set were obtained for BNN and Support Vector Machines (SVM) models, respectively.

From all the models under consideration, the BNN classification model had the highest prediction accuracy of detecting pits in cherries. To illustrate how cases belonging to different classes were classified by this model, the confusion matrices with distinction between the three cultivars studied were presented in Table 4. In the confusion matrices,

the diagonal cells show the number of residue positions that were correctly classified for each structural class, while the off-diagonal cells represent the number of residue positions that were misclassified. The results obtained for both the PCA-BNN and BNN (Table 4) models, as based on 27 selected wavelengths, were similar. In both models, the lowest number of misclassifications were observed for the 'Groniasta' and 'Łutówka' cultivars. The cherries with pit fragments were most frequently confused with

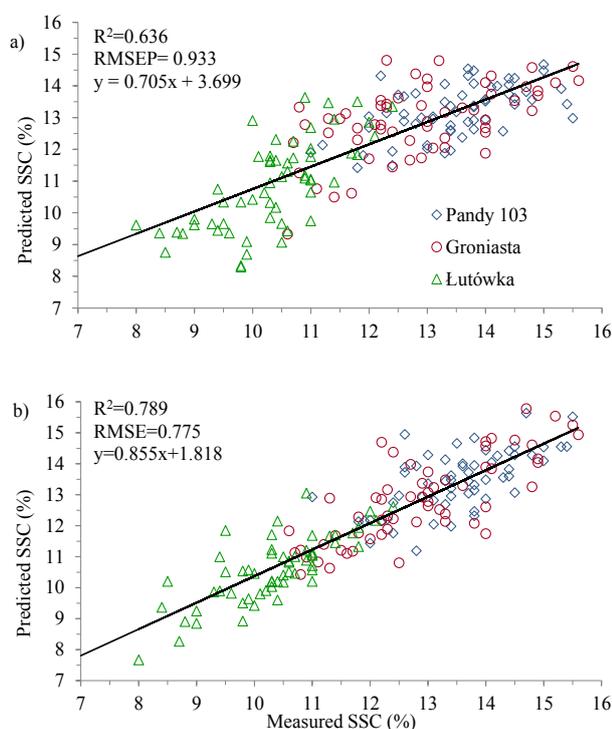


Fig. 4. Scatter plots of BNN predictions of soluble solid SSC versus refractometric measurements for the ‘Pandy 103’, ‘Groniasta’, and ‘Łutówka’ cherries. SSC predictions were obtained using the 5 first principal components (a) and 27 selected wavelengths, based on second derivatives of mean transmittance spectra (b).

cherries with a whole pit. These results suggest that the overall accuracy of distinguishing between pitted and intact cherries was very high (higher than 96%) and comparable to the classification precision (96.5%) of the whole pit detection by Qin and Lu (2005). It was also better than those (84.6% correctly classified instances) obtained by Xing *et al.* (2008) for detecting internal insect infestation in tart cherries. Moreover, our results are close to those obtained by Haff *et al.* (2013). They achieved recognition rates of 97.3% for the pitted and 94 % for un-pitted cherries using a one dimensional x-ray detector array. Of note, the same authors reported that x-ray imaging equipment is expensive and bulky, and its implementation on the processing line is cumbersome.

Because the cultivars were selected to cover the broad range of SSC variations, standard destructive tests were performed to verify whether the fruit material would be appropriate for modelling the SSC values on the basis of

hyperspectral measurements. The SSC statistics of the cherry samples of all three cultivars studied are summarized in Table 5. The SSC measurements of 180 samples belonging to the three cherry cultivars (60 samples each) were fairly normally distributed around the mean value (13.5, 12.6, and 10.6%, with standard deviation of 0.88, 1.01, and 1.66 for the ‘Pandy 103’, ‘Groniasta’, and ‘Łutówka’ cultivars, respectively). The normality of the SSC distribution was then assessed *via* the Shapiro-Wilk normality test at a significance level equal to 0.05. For the three cultivars studied, the W values ranged from 0.983 to 0.987. Taking into account that the critical value for 60 samples for this test is equal to 0.954 (smaller than all the obtained W values), the null hypothesis that the population is normally distributed cannot be rejected. It was expected that the differences in SSC between cultivars had an effect on the transmittance spectral characteristics of cherries. As can be seen in Table 4, the cultivar with the highest value of SSC (‘Pandy 103’) was characterized by the lowest classification accuracy when compared with the other studied cultivars. The high value of SSC has influence on the tissue optical density and, eventually, on the capability of pit detection.

To predict SSC on the basis of the hyperspectral transmittance data processed, soft computing modelling was performed *via* both PCA-BNN and BNN models. The modelling was carried out using second derivative spectra in the wavelength range of 600-1000 nm. In this study, 180 samples (60 samples for each cultivar) were divided into calibration and prediction sets (160:20), which were the same for both BNN and PCA-BNN models. The results obtained by modelling were subsequently compared with the measured SSC values obtained by destructive tests. To develop a reliable model, the range of data sets used for model creation and model validation should be independent. Another requirement is that the range of values should be relatively high in both sets. In this study, the variation range of the data set used for model creation was from 8.0-15.6%, and the variation range of the validation set was from 8.4-14.9%.

The relationships between the measured versus predicted values for SSC (%) including data of the three cherry cultivars studied, as obtained *via* the PCA-BNN and BNN models, are shown in Fig. 4a, b, respectively. The solid lines are the regression lines corresponding to the ideal correlation between the measured and predicted values. As evidenced, the predicted values are clustered

Table 5. The statistics of cherry sample SSC (number of samples 60)

| Cultivar | SSC (%) | SD (%) | CV | Kurtosis | Skewness | W |
|-------------|---------|--------|-------|----------|----------|-------|
| ‘Pandy 103’ | 13.5 | ±0.88 | 6.54 | -0.49 | -0.006 | 0.985 |
| ‘Groniasta’ | 12.6 | ±1.66 | 13.12 | -0.05 | -0.317 | 0.983 |
| ‘Łutówka’ | 10.6 | ±1.01 | 9.51 | -0.04 | 0.045 | 0.987 |

in two regions corresponding to ‘Łutówka’ and ‘Pandy 103’/‘Groniasta’ cultivars. Thus, a considerable difference in SSC exists between these two groups of data. Moreover, the minor differences in SSC content between ‘Pandy 103’ and ‘Groniasta’ brought about a situation in which the predicted values for these two cultivars could not be separated.

The prediction results showed that both models under consideration could be useful to estimate SSC. The prediction results for the PCA-BNN model are characterized by $R^2=0.636$ and $RMSEP=0.933\%$. The better prediction results of $R^2=0.789$ and $RMSEP=0.775\%$ were obtained *via* the BNN model with 27 selected wavelengths.

The results obtained in this study are comparable and, in many cases, superior to the previous attempts at predicting the SSC of fruits on the basis of hyperspectral imaging. The prediction accuracy of the SSC obtained in this research is better than that obtained by Leive-Valenzuela *et al.* (2014) for blueberries using transmission images ($R^2=0.52$ and $RMSEP=1.39\%$) and by Lu (2007) for apples based on the hyperspectral scattering technique ($R^2=0.76$ and $SEP=0.72\%$). Moreover, the second derivative treatment of transmission spectra gave better classification results ($R^2=0.789$ and $RMSEP=0.775\%$) than the non-derivatized spectra obtained by Herrera *et al.* (2003) for predicting the SSC of Chardonnay grapes ($R^2=0.777$ and $RMSEP=1.16\%$). Nonetheless, it is difficult to compare prediction accuracy when different species are studied and different methodologies are used. For instance, Sun *et al.* (2009) obtained a high SSC prediction accuracy for pears ($R^2=0.92$ and $RMSEP=0.53\%$) and Fan *et al.* (2009) did so for apples ($R^2=0.953$ and $SEP=0.384\%$). Similarly, Liu *et al.* (2010) obtained promising results of SSC prediction ($R=0.9$ and $RMSEP=0.68\%$ Brix) for Nanfeng mandarin and navel orange by using a combination PCA-BNN method with a multiplicative scatter correction (MSC) spectral pre-treatment. Cherries were studied by Carlini *et al.* (2000), with the SSC prediction accuracy proving very high ($R^2=0.97$ and $SEP=0.49\%$). Previous research has indicated that SSC prediction accuracy can depend on measurement mode (transmission, reflectance and inter-radiance). Schaare and Fraser (2000) assumed that inter-radiance mode spectra provided a very good prediction of internal properties of fruit. The SSC of kiwifruit, for example, in inter-radiance mode was predicted with $R^2=0.93$ and $SEP=0.8\%$ Brix, whereas in transmittance mode, $R^2=0.89$ and $SEP=1.01\%$ Brix.

CONCLUSIONS

1. On the basis of the results presented, it may be concluded that the proposed application of hyperspectral transmittance imaging increases the potential of pits or pit fragments detection in fresh cherries.

2. Two methods of spectral data reduction, namely the analysis of the maxima and minima on the pre-treated second derivatives of the spectra or by PCA, and the supervised classification methods, enabled the obtaining of very high classification accuracies (higher than 96%), when compared to earlier studies.

3. The applied methodology of hyperspectral image analysis gave high accuracy of whole pits and pit fragments detection within cherries for all variants of an experiment which covered a broad range of cherry sizes and SSCs. However, this accuracy has proven to be negatively correlated with SSC.

4. The unquestionable advantage of the solution presented is that the algorithms in question are based on a limited number of wavebands from the VNIR spectrum (27 wavelengths). This considerably increases the speed of the analysis. An effective alternative is the use of 5 PCA components when these are derived from the spectral cube being the input to the classification models.

5. The specific spectral characteristics in the transmission mode of cherries reflected the differentiation of their soluble solid content. This was made evident *via* back propagation neural networks models, by high correlation coefficients (up to 0.79) between the measured and predicted values.

6. Taking into account the technological development of the current hyperspectral sensors, manifested by the increasing speed of image registration and processing, it can be expected that the adaptation of the presented solution on the sorting lines is quite realistic.

Conflict of interest: The Authors do not declare conflict of interest.

REFERENCES

- Allen K.M., Vandyke L.H., and Brunsbach R.L., 1966. Apparatus for detecting seeds in fruit. U.S. Patent No. 3, 275-136.
- Baiano A., Terracone C., Peri G., and Romaniello R., 2012. Application of hyperspectral imaging for prediction of physico-chemical and sensory characteristics of table grapes. *Computers Electronics Agric.*, 87, 142-151.
- Baranowski P., Jedryczka M., Mazurek W., Babula-Skowronska D., Siedliska A., and Kaczmarek J., 2015. Hyperspectral and thermal imaging of oilseed rape (*Brassica napus*) response to fungal species of the genus *Alternaria*. *PLoS one*, 10(3), e0122913.
- Baranowski P., Mazurek W., and Pastuszka-Woźniak J., 2013. Supervised classification of bruised apples with respect to the time after bruising on the basis of hyperspectral imaging data. *Postharvest Biol. Technol.*, 86, 249-258.
- Becker B.L., Lusch D.P., and Qi J., 2005. Identifying optimal spectral bands from in situ measurements of Great Lakes coastal wetlands using second-derivative analysis. *Remote Sensing Environ.*, 97, 238-248.
- Carlini P., Massantini R., and Mencarelli F., 2000. Vis-NIR measurement of soluble solids in cherry and apricot by PLS regression and wavelength selection. *J. Agric. Food Chemistry*, 48, 5236-5242.

- Donis-González I.R., Guyer D.E., Kavdir I., Shahriari D., and Pease A., 2015.** Development and applicability of an agarose-based tart cherry phantom for computer tomography imaging. *J. Food Measurement Characterization*, 9(3), 290-298.
- Fan G., Zha J., Du R., and Gao L., 2009.** Determination of soluble solids and firmness of apples by Vis/NIR transmittance. *J. Food Eng.*, 93 (4), 416-420.
- Fan S., Huang W., Guo Z., Zhang B., and Zhao C., 2005.** Prediction of soluble solids content and firmness of pears using hyperspectral reflectance imaging. *Food Anal. Method*, 8, 1936-1946.
- FAOSTAT, 2016.** <http://wwandw.fao.org/faostat/en/?#data/QC> Accessed 5.01.2017.
- Golic M. and Walsh K.B., 2006.** Robustness of calibration models based on near infrared spectroscopy for the in-line grading of stone fruit for total soluble solids content. *Analytica Chimica Acta*, 555(2), 286-291.
- Haff R.P., Jackson E.S., and Pearson T.C., 2005.** Non-destructive detection of pits in dried plums. *Small*, 175, 6-6.
- Haff R., Pearson T., and Jackson E., 2013.** One dimensional linescan x-ray detection of pits in fresh cherries. *American J. Agric. Sci. Technol.*, 1, 18-26.
- Haff R.P. and Toyofuku N., 2008.** X-ray detection of defects and contaminants in the food industry. *Sensing Instrumentation for Food Quality and Safety*, 2 (4), 262-273.
- Herrera J., Guesalaga A., and Agosin E., 2003.** Shortwave-near infrared spectroscopy for non-destructive determination of maturity of wine grapes. *Measurement Sci. Technol.*, 14(5), 689.
- Kawano S., 2016.** Past, present and future near infrared spectroscopy applications for fruit and vegetables. *NIR news*, 27(1), 7-9.
- Leiva-Valenzuela G.A, Lu R., and Aguilera J.M., 2014.** Assessment of internal quality of blueberries using hyperspectral transmittance and reflectance images with whole spectra or selected wavelengths. *Innovative Food Sci. Emerg. Technol.*, 24, 2-13.
- Liaghat S., Ehsani R., Manso S., Shafr H.Z., Meon S., Sankaran S., and Azam S.H., 2014.** Early detection of basal stem rot disease (Ganoderma) in oil palms based on hyperspectral reflectance data using pattern recognition algorithms. *Int. J. Remote Sensing*, 35(10), 3427-3439.
- Liu Y., Sun X., and Aiguo O., 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.*, 44 (4), 602-607.
- Lu R., 2007.** Nondestructive measurement of firmness and soluble solids content for apple fruit using hyperspectral scattering images. *Sensing and Instrumentation for Food Quality and Safety*, 1(1), 19.
- Lu Y., Huang Y., and Lu R., 2017.** Innovative hyperspectral imaging-based techniques for quality evaluation of fruits and vegetables: A Review. *Applied Sci.*, 7(2), 189.
- Munera S., Besada C., Aleixos N., Talens P., Salvador A., Sun D.W., Cubero S., and Blasco J., 2017.** Non-destructive assessment of the internal quality of intact persimmon using colour and VIS/NIR hyperspectral imaging. *LWT-Food Sci. Technol.*, 77, 241-248.
- Nicolai B.M., Defraeye T., De Ketelaere B., Herremans E., Hertog M.L., Saeys W., Toricelli A., Vandendriessche T. and Verboven P., 2014.** Nondestructive measurement of fruit and vegetable quality. *Annual Review Food Sci. Technol.*, 5, 285-312.
- Pan L., Sun Y., Xiao H., Gu X., Hu P., Wei Y., and Tu K., 2017.** Hyperspectral imaging with different illumination patterns for the hollowness classification of white radish. *Postharvest Biology Technol.*, 126, 40-49.
- Pu Y.Y., Feng Y.Z., and Sun D.W., 2015.** Recent progress of hyperspectral imaging on quality and safety inspection of fruits and vegetables: a review. *Comprehensive Reviews in Food Sci. Food Safety*, 14(2), 176-188.
- Qin J. and Lu R., 2005.** Detection of pits in tart cherries by hyperspectral transmission imaging. *Trans. ASAE*, 48 (5), 1700-1963.
- Schaare P.N. and Fraser D.G., 2000.** Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (*Actinidiachinensis*). *Postharvest Biol. Technol.*, 20(2), 175-184.
- Siedliška A., Baranowski P., and Mazurek W., 2014.** Classification models of bruise and cultivar detection on the basis of hyperspectral imaging data. *Computers Electronics Agric.*, 106, 66-74.
- Song D., Song L., Sun Y., Hu P., Tu K., Pan L., Yang H., and Huang M., 2016.** Black Heart Detection in White Radish by Hyperspectral Transmittance Imaging Combined with Chemometric Analysis and a Successive Projections Algorithm. *Applied Sci.*, 6(9), 249.
- Sun J., Ma B., Dong J., Zhu R., Zhang R., and Jiang W., 2016.** Detection of internal qualities of hami melons using hyperspectral imaging technology based on variable selection algorithms. *J. Food Process Eng.*, 40, doi: 10.1111/jfpe.12496
- Sun T., Lin H., Xu H., and Ying Y., 2009.** Effect of fruit moving speed on predicting soluble solids content of 'Cuiguan' pears (*Pomaceae pyrifolia* Nakai cv. Cuiguan) using PLS and LS-SVM regression. *Postharvest Biol. Technol.*, 51(1), 86-90.
- Sun Y., Gu X., Sun K., Hu H., Xu M., Wang Z., Kang T., and Pan L., 2017.** Hyperspectral reflectance imaging combined with chemometrics and successive projections algorithm for chilling injury classification in peaches. *LWT-Food Sci. Technol.*, 75, 557-564.
- Szuvandzsie P., Helyes L., Lugasi A., Szántó C., Baranowski P., and Pék Z., 2014.** Estimation of antioxidant components of tomato using VIS-NIR reflectance data by handheld portable spectrometer. *Int. Agrophys.*, 28(4), 521-527.
- Timm E.J., Gilliland P.V., Brown G.K., and Affeldt H.A., 1991.** Potential methods for detecting pits in tart cherries. *Applied Eng. Agric.*, 7(1), 103-109.
- USDA foreign agricultural service - EU-28, 2016.** Stone Fruit Annual, https://gain.fas.usda.gov/Recent%20GAIN%20Publications/Stone%20Fruit%20Annual_Madrid_EU-28_8-19-2016.pdf
- Wang N.N., Sun D.W., Yang Y.C., Pu H., and Zhu Z., 2016.** Recent Advances in the Application of Hyperspectral Imaging for Evaluating Fruit Quality. *Food Analytical Methods*, 9(1), 178-191.
- Williams P. and Norris K. (Eds), 2001.** Near-Infrared Technology in the Agricultural and Food Industries. American Association of Cereal Chemists, St. Paul, MIN, USA.

- Witten I.H. and Frank E., 2005.** Data mining. Practical machine learning tools and techniques. Morgan Kaufmann.
- Wojdylo A., Nowicka P., Laskowski P., and Oszmiański J., 2014.** Evaluation of sour cherry (*Prunuscerasus L.*) fruits for their polyphenol content, antioxidant properties, and nutritional components. *J. Agric. Food Chemistry*, 62(51), 12332-12345.
- Xing J., Guyer D., Ariana D., and Lu R., 2008.** Determining optimal wavebands using genetic algorithm for detection of internal insect infestation in tart cherry. *Sensing and Instrumentation for Food Quality and Safety*, 2(3), 161-167.
- Zion B., Kim S.M., McCarthy M.J., and Chen P., 1997.** Detection of pits in olives under motion by nuclear magnetic resonance. *J. Sci. Food Agric.*, 75(4), 496-502.
- Zion B., McCarthy M.J., and Chen P., 1994.** Real-time detection of pits in processed cherries by magnetic resonance projections. *LWT-Food Sci. Technol.*, 27(5), 457-462.