Use of Fourier Transform Infrared Spectroscopy and Chemometric Data Analysis To Evaluate Damage and Age in Mushrooms (Agaricus bisporus) Grown in Ireland

Aoife O’Gorman
*Technological University Dublin*, aoife.ogorman@tudublin.ie

Gerard Downey
*Teagasc*

Aoife Gowen
*University College Dublin*, aoife.gowen@ucd.ie

Catherine Barry-Ryan
*Technological University Dublin*, Catherine.Barryryan@tudublin.ie

Follow this and additional works at: [https://arrow.tudublin.ie/schfsehart](https://arrow.tudublin.ie/schfsehart)

**Part of the Food Chemistry Commons**

**Recommended Citation**

Use of Fourier-transform infrared spectroscopy and chemometric data analysis to evaluate damage and age in mushrooms (*Agaricus bisporus*) grown in Ireland

Aoife O’Gorman⁵, Gerard Downey⁶, Aoife A. Gowen⁷, Catherine Barry-Ryan⁵, and Jesus M. Frias⁵⁺

⁵ School of Food Science & Environmental Health, Dublin Institute of Technology, Cathal Brugha Street, Dublin 1, Ireland
⁶ Teagasc, Ashtown Food Research Centre, Ashtown, Dublin 15, Ireland
⁷ Biosystems Engineering, School of Agriculture, Food Science and Veterinary Medicine, University College Dublin, Dublin 4, Ireland

⁺ Corresponding author. Tel +353 1 402 4459, Fax +353 1 402 4495
E-mail: jesus.frias@dit.ie
Abstract

The aim of this research was to investigate whether the chemical changes induced by mechanical damage and aging of mushrooms can be (a) detected in the mid-infrared absorption region and (b) identified using chemometric data analysis. Mushrooms grown under controlled conditions were bruise-damaged by vibration to simulate damage during normal transportation. Damaged and non-damaged mushrooms were stored for up to 7 days post-harvest. Principal component analysis of FTIR spectra showed evidence that physical damage had an effect on tissue structure and the aging process. Random forest classification models were used to predict damage in mushrooms producing models with error rates of 5.9 and 9.8% with specific wavenumbers identified as important variables for identifying damage, PLS models were developed producing models with low levels of misclassification. Modeling post-harvest age in mushrooms using random forests and PLS resulted with high error rates and misclassification; however, random forest models had the ability to correctly classify 82% of day zero samples, which may be a useful tool in discriminating between ‘fresh’ and old mushrooms. This study highlights the usefulness of FTIR spectroscopy coupled with chemometric data analysis in particular for evaluating damage in mushrooms and with the possibility of developing a monitoring system for damaged mushrooms using the FTIR ‘fingerprint’ region.

Key words: FTIR spectroscopy; chemometrics; mushrooms; aging; damage
Mushroom cultivation is a worldwide business with the global market valued at over $45 billion in 2005 \((1)\). In Ireland more than 60,000 tons of button mushrooms \((Agaricus bisporus)\) are produced annually, making them one of the most important horticultural crops grown \((2)\). Mushrooms are one of the most perishable food products with a maximum shelf-life of 3-4 days at ambient temperature \((3)\) mainly because they have no cuticle to offer protection from physical damage, microbial attack or water loss \((4)\). They may be bruised easily by physical stress during harvesting, handling and transportation. This mechanical damage triggers a browning process which is the major cause of loss of value in the market \((5, 6)\). A second significant factor determining mushroom quality is time elapsed between harvesting and delivery to the marketplace. Post-harvest age is particularly important for any mushroom exporting country \(i.e.\) Ireland for which access to the food markets in larger, neighboring countries within Europe is vital. There is a need for a method which would allow objective evaluation of mushroom quality to ensure that only high quality produce reaches the retail market and that is able to produce information on the metabolites in mushrooms affected by senescence and damage \((7)\).

Fourier-transform infrared spectroscopy is an analytical technique that enables the rapid, reagentless and high-throughput analysis of a diverse range of samples \((8)\). Its importance lies in its ability to allow rapid and simultaneous characterization of different functional groups such as lipids, proteins, nucleic acids and polysaccharides \((9-12)\) in biological molecules and complex structures. FTIR spectroscopy is an important tool used for quality control and process monitoring in the food industry because it is less expensive, has better performance and is easier to use than other
methods (13). In the same way, FTIR spectroscopy has been used as a fingerprinting tool to study response of cells to various stressing situations (14-16).

A key to the successful operation of this technique is the availability of mathematical tools for the interrogation and mining of large spectral data sets. Principal component analysis (PCA), partial least squares (PLS) regression and random forests (RF) are chemometric tools that have been successfully used to extract information from FTIR data (17, 18).

The objective of this study was to investigate the damage and aging of mushrooms grown in Ireland using FTIR spectroscopy in order to (a) differentiate between damaged and undamaged mushrooms and (b) to determine mushroom post-harvest age. The ability to develop a tool that could detect physical damage before browning becomes visible would be of importance to the mushroom industry and could reduce economic losses.

MATERIALS AND METHODS

Mushrooms. Second flush mushrooms were grown at the Teagasc Research Centre Kinsealy (Dublin, Ireland), harvested damage-free. A set of 160 closed cap, defect-free Agaricus bisporus strain Sylvan A15 (Sylvan Spawn Ltd., Peterborough, United Kingdom) mushrooms (3-5 cm cap diameter) were selected for this study and immediately transported by road to the testing laboratory. Special trays were designed to hold mushrooms by the stem using a metal grid to avoid contact between (a) mushrooms and (b) between the top of mushroom caps and the tray lid during transportation. Mushrooms arrived at the laboratory premises within 1 h after harvesting and where either damaged for the specified time length or remained damage free and then stored at 4ºC until required for analysis.

Mushroom treatments. Mushrooms (n=160) were harvested in the conventional
manner on a single occasion. On the day of harvest, a subset (n=80) was subjected to
damage using a mechanical shaker (Gyrotory G2, New Brunswick Scientific
Co. USA) set at 300 rpm (rotations per minute) for 20 minutes; these samples were
labeled as damaged (D). The remaining 80 mushrooms were untreated and labeled
undamaged (UD). Ten (10) damaged and 10 undamaged mushrooms were selected at
random from their respective sub-sets on the day of harvesting and prepared for
spectroscopic analysis (see below); these are referred to as day 0 samples. The
remainder of the mushrooms (70 each of damaged and undamaged) were placed in
plastic punnets (six mushrooms per punnet) and stored as separate batches at 4ºC in a
controlled temperature facility. On each of 7 consecutive days of such storage, a set
of 10 damaged and 10 undamaged mushrooms was randomly selected, removed from
storage and prepared for FTIR analysis.

**FTIR spectroscopy.** Sample preparation involved the manual dissection of each
mushroom into its three main tissue types (cap, gills and stalk) before freezing
overnight at -70ºC in a cryogenic refrigerator (Polar 340V: Angelantoni Industrie
spA, Massa Martana, Italy) followed by freeze-drying (Micro-modulyo, EC
Apparatus Inc, New York, USA) for 24 h. Freeze-dried samples were manually
ground into fine particles using a pestle and mortar. Then, 9 mg (3% w/w) of each
sample was mixed with 291 mg (97% w/w) KBr (Sigma Aldrich, Dublin, Ireland).
KBr pellets were prepared by exerting pressure of 100 kg/cm² (1200 psi) for
approximately 2 min in a pellet press (Specac, UK). To eliminate any interference
which might be caused by variation in pellet thickness different pellets were prepared
from the same sample and their infrared spectra compared. These samples were
identical with their average spectra used for analysis (19).
Spectra were collected using a Nicolet Avatar 360 FTIR E.S.P (Thermo Scientific, Waltham, MA, USA) over the frequency range 4000-400 cm\(^{-1}\). One hundred scans of each pellet were collected at 4 cm\(^{-1}\) resolution at room temperature using OMNIC software (version ESP 5.1). The average of the 100 scans was used for further data analysis. FTIR spectral data were discretized resulting in spectra containing 1868 individual points (discretised every 2 cm\(^{-1}\)) for chemometric analysis.

**Chemometric data analyses.** Multivariate models for damage and age prediction in mushrooms using both raw (i.e. unmodified) and pre-treated spectral data were developed; the pre-treatment used was standard normal variate [SNV] and was intended to reduce scatter-induced effects in the spectra (20). The frequency region studied was 2000-400 cm\(^{-1}\) (fingerprint region); this spectral range encompasses absorptions from most of the chemical species present and attenuation of the dataset in this way avoids spectral regions which have low information content and may therefore interfere with effective model development.

Random forest modeling achieves a classification by constructing a series of decision trees (21) and takes input variables down all trees in order to optimize classification. Each tree is constructed using a different bootstrap sample from the original data, about one-third of the cases are left out of the bootstrap sample and are not used in the construction of the k-th tree. These sets of unseen samples are called out-of-bag (OOB) sets. RF makes use of these OOB sets in many ways, in particular to give an unbiased estimate of the prediction error on unseen cases (22).

Random forest models were built to (a) discriminate between damaged and undamaged mushrooms and (b) to predict mushroom ages. The number of trees fitted to build the random forest was 1000, the number of random wavenumbers tried at every node of the tree was set at 500 after optimization and the random forest model
trained was made using a stratified random sampling strategy of the sample spectra that would take the same number of samples from each of the tissues. Principal component analysis (PCA) was used to identify patterns in data in a way which emphasizes differences and similarities. It is used to indicate relationships among groups of variables in a data set and show relationships that might exist between objects (23).

Partial least squares (PLS) regression was applied to the spectral data sets to develop a quantitative model for prediction of the age of damaged mushrooms. A common problem in development of multivariate prediction models is selection of the optimum number of PLS loadings; often, this selection is based on an examination of the RMSECV but identification of a minimum is not always possible or unambiguous and sub-optimal models incur a significant risk of overfitting. Experience has shown that this can be a problem when parameters which are of practical relevance, such as post-harvest age or damage, but have unclear molecular basis are being modeled. In order to avoid overfitting, model cross validation was employed as follows:

1. Samples were randomly-designated from each tissue/damage status/time grouping as calibration (60%) or validation (40%) samples. The validation subset was left completely out during the optimization of model based on the calibration set.

2. The model optimization step was carried out in order to estimate the optimal dimensionality of the PLS model built on the calibration set. The method employed for this was based on the observation that an indication of overfitting is the appearance of noise in regression vectors; this takes the form of a reduction in apparent structure and the presence of sharp peaks with a high degree of directional oscillation. A simple method (24) for objectivity
quantifying the shape of a regression vector, combined with the root mean square error of cross-validation (RMSECV) for the calibration set was applied in this study.

3. The random sample designation, model development and evaluation were performed 100 times. At the end of this cycle, models were initially examined on the basis of the number of latent variables selected, the most common number was then chosen as the optimum.

Mushroom discrimination (damaged versus undamaged) were performed using partial least squares discrimination analysis (PLS-DA). For PLS-DA, a dummy Y-variable was assigned to each mushroom tissue sample, 1 for damaged and 0 for undamaged. PLS-DA calibration models were developed and assessed using 100 randomly-populated calibration and validation sample sets.

Principal component analysis (PCA) and partial least squares (PLS) regression were performed using MATLAB and The Unscrambler software (v.9.7; Camo A/S, Oslo, Norway). The routine for selection of the optimum number of PLS loadings was also performed in MATLAB. Random forest modeling was performed using R 2.8.0 (25).

RESULTS AND DISCUSSION

Spectral data. Average raw spectra of each of the three tissue types collected from all the damaged and undamaged samples (day 0-7 in each case) are shown in Figure 1(a, b and c). A number of observations may be made on these spectra. Firstly, the major feature is a vertical offset from one average plot to another; this offset originates in light scatter effects and may be a complication in further data analysis. Average spectra of the three tissue types also bear a close resemblance to
each other; there is little visible difference in peak minima locations in Figure 1. In
terms of minima locations, there are major bands at 1650, 1090, 1020, 935 cm\(^{-1}\);
minor minima may be seen at 1560, 1150 and 1050 cm\(^{-1}\) (Figure 2). Unambiguous
identification of the molecular source of features in mid-infrared spectra of biological
material is difficult but the peak at 1650 may be attributed to an amide I group while
at 1560 cm\(^{-1}\) may be identified as resulting from amide II groups (26, 27). Both major
absorbance peaks at 1090 and 1020 cm\(^{-1}\) have been attributed as structures in chitin, a
major structural polysaccharide in mushrooms; absorbance at 1090 cm\(^{-1}\) may also
arise from secondary alcohols. Smaller features at 1150 and 1050 cm\(^{-1}\) have been
attributed to tertiary and primary alcohol structures (28). Minima at 935, 890 and 874
cm\(^{-1}\) bands correspond to \(\alpha\)- or \(\beta\)- anomer C\(_1\)-H deformations. The bands at 935 and
890 cm\(^{-1}\) are attributed to glucan bands, while the band at 874 cm\(^{-1}\) is assigned to a
mannan band (29-31). An inability to attribute all spectral features is a common
feature of spectroscopy of complex biological matrices but the presence of such
spectral detail implies the detection of a significant quantity of information which
may be usefully interrogated by multivariate mathematical methods.

**Principal component analysis (PCA).** Undamaged samples were studied
separately on the basis of their tissue type i.e. caps, gills and stalks. Initial PCA of the
mushroom caps data revealed a single sample (day 7) which lay anomalously at some
significant distance from the others; this was deleted and the resulting score plot is
shown in Figure 3 for PC1 vs PC2; these first two principal components accounted for
97 and 2% respectively of the total variance in the spectral dataset and some sample
clustering on the basis of storage time is readily apparent. As a general observation, it
may be stated that the majority of the day 0 mushroom caps have a score value on
PC1 greater than zero and are therefore located on the right-hand-side of Figure 3a.
While there are indications that samples of different storage time cluster together, the spread of these clusters is quite large and it is not possible to readily discern any trend relating plot position and storage time in the plots. There is a suggestion that the dispersion of the samples decreases as the length of storage time increases. With regard to undamaged gill tissue, observations similar to those made above in relation to undamaged caps may be made although the distribution patterns are somewhat different.

In the case of damaged mushroom tissues, a different pattern was found. It is clear from Figure 3c, d and e that day 0 samples clustered together but separately from those of day 1 to day 7 samples irrespective of tissue type. This strongly suggested that physical damage had a significant effect on tissue structure and the subsequent aging process. Some implications regarding the rate of change of mushroom tissue composition with aging may be garnered from the observation that separation of day 0 from all other subsequent days accounts for the most variation in the spectral collection of damaged mushroom caps, gills and stalks.

Examination of PC loadings may provide information on the absorbing species which are involved in separations observed on a PC scores plot; however, meaningful interpretation of loadings arising from this dataset (data not shown) was not possible.

Detection of damage (random forests). The first random forest model developed attempted to identify which wavenumbers could be used to predict damage specifically. The model tried to predict damage in mushrooms using the IR spectra, a variable indicating the tissue from which the spectra originated (cap, gill or stalk) and the age of the mushroom (in days from 0-7) as explanatory variables. This resulted in good classification between damaged and undamaged samples with an out-of-bag error rate (OOB) of 5.9%, sensitivity of 93.3% and specificity of 95%.
In random forests there are two measures of importance to indicate how informative a particular variable (a wavenumber in our case) is, the mean decrease in accuracy and the Gini index. The decrease in Gini index is not as reliable as the marginal decrease in accuracy (32, 33) and for that reason the latter was analysed. The variables containing the most importance for predicting damage in the model are shown in Figure 4a. The most important variable for predicting damage was the age of the mushrooms followed by the wavenumbers 1868, 1870 and 1845 cm\(^{-1}\).

Induced damage in mushrooms leads to an enzymic response which is followed by brown discoloration. The enzymes involved in this response, tyrosinase, or polyphenol oxidases, catalyse the oxidation of phenols, which in turn promote the formation of melanin like compounds. This reaction is found not only in damaged mushrooms, but is also part of the natural aging process, with color in mushrooms becoming darker and less firm during storage (34). The three wavenumbers identified have the ability to differentiate between the chemical changes that are induced by the mechanical damage and are independent of those that take place due solely to aging. The three wavenumbers identified above are unassigned peaks.

By removing the variable age from the model a second model was built which would take IR spectra of mushrooms (independently of their age) and try to predict whether there is damage or not. This random forest could be used as a classifier of mushroom damage and gave a very good prediction model with an OOB error rate of 9.8%, sensitivity of 89.2% and specificity of 91.2%. Even receiving mushrooms whose storage time after harvest was unknown the model would still classify damaged and undamaged mushroom samples with a very good classification rate. The variables of importance involved in this classification model are shown in Figure 4b.
The most important variable for predicting damage according to the mean decrease accuracy plot is tissue used in the analysis followed by the wavenumbers 1868, 1870 and 1560 cm\(^{-1}\). The peak at 1560 cm\(^{-1}\) is attributed to amide II vibrations of proteins (29). Amide II bands along with amide I bands are major regions of the protein infrared spectrum. Amide II bands are associated with an out-of-phase combination of in-plane C-N stretching and N-H bending of amide groups (35). Absorption of this band was found to be higher in damaged samples and therefore an important variable for detecting damage in mushroom samples. The wavenumbers 1868 and 1870 cm\(^{-1}\) are unassigned.

**Detection of damage (PLS).** PLS-DA models were developed to discriminate between undamaged and damaged mushrooms of all tissue types separately. A summary of the average and dispersion of the results obtained on a percentage basis for each tissue is shown in Table 2; it is apparent that misclassification errors associated with all models were low, especially so in the case of gills and stalks. In terms of numbers of samples misclassified, these percentages translate to 1 or 2 only in each case. These results indicate that FTIR of freeze-dried mushroom tissues (especially gills and stalks) may be used to discriminate between damaged and undamaged mushrooms aged post-harvest from 0 to 7 days with almost complete confidence.

Modeling damage in mushrooms has been reported in literature in 2008 by Gowen and colleagues and in 2009 by Esquerre et al. (36, 37). Gowen and colleagues investigated the use of hyperspectral imaging and principal components analysis (PCA) to develop models to predict damage on mushroom caps with correct classification ranging from 79-100%. Using near infra-red spectroscopy and partial least squares (PLS) regression, Esquerre and colleagues were able to correctly classify
undamaged mushrooms from damaged ones with an overall correct classification model with 99% accuracy. The models for predicting damage using FTIR and random forests correctly classified 94 and 90% of samples respectively, whilst the PLS predictive models correctly classified 92-99% of undamaged samples from damaged ones. These results highlight the usefulness of FTIR and chemometrics for detecting physical damage in mushrooms with the possibility of developing a classification system for the industry.

**Predicting post-harvest age (random forests).** Initial random forest models were built to try and predict the mushroom age from day zero to day seven (0-7) using the IR spectra from the tissues and knowing whether they had been subjected to damage or not with the aim to identify specific wavenumbers associated with aging. The random forest model produced an OOB error rate of 32% i.e. 68% of samples were correctly classified. The results of the model fit are shown in Table 3. Misclassification of samples was seen for all mushroom ages particularly days 4, 5 and 7. Classification of day zero samples performed quite well in the model with 82% of samples correctly classified, which leads to the possibility of using IR spectroscopy as a tool to discriminate fresh mushrooms (D0) from mushrooms that have been subjected to refrigeration. This type of tool could enable packers and producers to avoid fraud and ‘recycling’ of product, supporting the evidence from visual inspection. The variables of importance identified by the mean decrease accuracy plot were damage, tissue type and the wavenumbers 399, 952 and 1508 cm\(^{-1}\).

A second model was developed to predict age using the same approach as above but removing the damage variable from the model. The model performed much the same as above with an OOB error rate of 33%; again misclassification within all sample ages was seen. The model correctly classified 79% of day zero models. The
important variables identified to predict age were tissue type and the wavenumbers 309, 952 and 1508 cm$^{-1}$. The peak at 952 cm$^{-1}$ is a glucan band ($\beta$-anomer C-H deformation) (29), glucans play many different roles in the physiology of fungi, some accumulate in the cytoplasm as storage, however most are present in the cell wall structure (38). This suggests that the ability to model aging in mushrooms may depend on the affect of glucan levels changing in the cell wall due to natural senescence. The wavenumbers at 399 and 1508 cm$^{-1}$ are unassigned. The OOB errors produced to model aging were quite large >33% which may be due to the low sample numbers.

**Predicting post-harvest age (PLS).** PLS regression was applied separately to the caps, gills and stalks datasets in an attempt to develop separate quantitative models for prediction of the age of mushrooms, both damaged and undamaged. Selection of the appropriate number of latent variables for each model was assessed on the basis of the frequency of their occurrence. As shown in Figure 5, this was a clear and unambiguous choice. A summary of the results obtained using mushrooms from day 0 to day 7 inclusive is shown in Table 3. In the case of undamaged mushrooms, root mean squared error of cross validation (RMSECV) values achieved were relatively high, only permitting the prediction of post-harvest age of damaged mushrooms to within ± 2 to 3 days approximately (95% confidence limit) depending on tissue type. The practical utility of such accuracy levels may be gauged by examination of the SD/RMSECV ratio, all but one of which are below 3.0, the generally accepted minimum value for a model to be of practical utility. With regard to damaged mushrooms, model predictive accuracies were similar for caps and stalks with RMSECV (and RER) values of 1.3 (1.9) and 1.2 (2.) respectively. In the case of gill tissue, better predictive accuracy was achieved with RMSECV and RER values equal
to 0.8 and 3.1 respectively. The number of latent variables associated with these models was low and similar in all cases, with a variation between 6 and 8 only. The application of an objective indicator of the optimum number of PLS loadings to include in any model contributed to their stable performance.

The results presented for modeling age in mushrooms using FTIR and chemometrics had misclassification errors of over 30% (random forests) yielding relatively unsuccessful results. However, random forest models were able to classify day zero samples reasonably well with correct classifications of 82 and 79% which leads to the possibility of using IR spectroscopy in detecting fresh mushrooms from old mushrooms and could be used within the sector for detecting fraud and ‘recycling’ of product. The time required for freeze-dried sample preparation is in the order of hours, thus this approach would be applicable for research and quality control purposes.

**Acknowledgments:**

The authors acknowledge financial support from the Irish Department of Agriculture and Food under the Food Institutional Research Measure (FIRM), supported through EU and national funds. Thanks are due to Dr. Helen Grogan and Ted Cormican, Teagasc, Kinsealy Research Centre, Dublin, Ireland for the supply of mushrooms and background information.


18. Rossel, R. A. V.; Behrens, T., Using data mining to model and interpret soil

19. Garip, S.; Gozen, A. C.; Severcan, F., Use of Fourier transform infrared
spectroscopy for rapid comparative analysis of Bacillus and Micrococcus isolates.

20. Preisner, O.; Lopes, J. A.; Menezes, J. C., Uncertainty assessment in FT-IR
spectroscopy based bacteria classification models. Chemometrics and Intelligent

Biology, 27th Annual Conference, Shanghai, China, 2005; Shanghai, China, 2005.

22. Granitto, P. M.; Gasperi, F.; Biasioli, F.; Trainotti, E.; Furlanello, C., Modern
data mining tools in descriptive sensory analysis: A case study with a Random forest

Elseheri, W., A chemometric study of chromatograms of tea extracts by correlation
optimization warping in conjunction with PCA, support vector machines and random

for estimation of PLS model dimensionality using regression vectors. Journal of
Chemometrics 2010, in press.

25. R_Development_Core_Team R: a language and environment for statistical

26. Belton, P. S.; Colquhoun, I. J.; Grant, A.; Wellner, N.; Field, J. M.; Shewry, P.
R.; Tatham, A. S., FTIR and NMR studies on the hydration of a high-Mr subunit of

27. Di Mario, F.; Rapanà, P.; Tomati, U.; Galli, E., Chitin and chitosan from

28. Workman, J., Methods and Interpretations in "Handbook of Organic

29. Mohacek-Grosev, V.; Bozac, R.; Puppels, G. J., Vibrational spectroscopic
characterization of wild growing mushrooms and toadstools. Spectrochimica Acta

30. Pierce, J. A.; Rast, D. M., A comparison of native and synthetic mushroom
melanins by fourier-transform infrared spectroscopy. Phytochemistry 1995, 39, (1),
49-55.

31. Sandula, J.; Kogan, G.; Kacurakova, M.; Machova, E., Microbial (1->3)-beta-
d-glucans, their preparation, physico-chemical characterization and

Floyd, E.; Zhao, H., Pathway analysis using random forests classification and


34. Eastwood, D.; Burton, K., Mushrooms - a matter of choice and spoiling

35. Militiello, V.; Casarino, C.; Emanuele, A.; Giostra, A.; Pullara, F.; Leone, M.,
Aggregation kinetics of bovine serum albumin studied by FTIR spectroscopy and

36. Gowen, A. A.; O'Donnell, C.; Taghizadeh, M.; Cullen, P. J.; Frias, J.;
Downey, G., Hyperspectral imaging combined with principle component analysis for


FIGURE CAPTIONS

Figure 1 FTIR transmittance spectra of all mushroom tissues in (a) 400-1800 cm⁻¹ (b) 2800-3050 cm⁻¹, and (c) 3050 – 4000 cm⁻¹ wavenumber ranges

Figure 2. Average undamaged caps spectrum (raw data)

Figure 3. PC1vs PC2 score plots of undamaged mushroom tissue (a) caps; (b) gills (c) stalks and damaged tissue (d) caps; (e) gills and (f) stalks

Figure 4(a) Relative importance plot of variables that are important in the random forest model for predicting damage/undamaged samples. The variable age being the most important followed by the wavenumbers 1868, 1870 and 1845 cm⁻¹. 4(b) Relative importance plot of variables that are important in the random forest model for predicting damaged/undamaged samples when age is not a variable. The most important variables are tissue type followed by the wavenumbers 1868, 1870 and 1560 cm⁻¹

Figure 5 Frequency of generation of PLS regression models for mushroom post-harvest age on the basis of the number of latent variables selected. (a) undamaged caps, (b) undamaged gills, (c) undamaged stalks, (d) damaged caps, (e) damaged gills and (f) damaged stalks. Abscissa – no. of latent variables in model; ordinate – number of occurrences

TABLES

Table 1 Summary of results for mushroom discrimination on the basis of damage

<table>
<thead>
<tr>
<th></th>
<th>#Samples</th>
<th>#Loadings</th>
<th>% undamaged misclassified mean (std. deviation)</th>
<th>% damaged misclassified mean (std. deviation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Caps</td>
<td>160</td>
<td>7</td>
<td>4.1 (4.3)</td>
<td>7.6 (4.0)</td>
</tr>
<tr>
<td>Gills</td>
<td>160</td>
<td>9</td>
<td>2.1 (3.0)</td>
<td>0.8 (1.7)</td>
</tr>
<tr>
<td>Stalks</td>
<td>160</td>
<td>12</td>
<td>1.7 (2.1)</td>
<td>0.6 (1.5)</td>
</tr>
</tbody>
</table>

Table 2 Confusion matrix and the error rate for the prediction of mushroom age. The OOB error rate: 32%. The highlighted numbers are correctly classified samples

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td><strong>49</strong></td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td><strong>42</strong></td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>5</td>
<td><strong>43</strong></td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td><strong>47</strong></td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td><strong>32</strong></td>
<td>2</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>12</td>
<td>3</td>
<td><strong>29</strong></td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td><strong>48</strong></td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>6</td>
<td>8</td>
<td><strong>34</strong></td>
</tr>
</tbody>
</table>

0-7: Sample age in days from day zero to day seven
Error rate: The % misclassification for each sample age
Table 3 Summary of PLS regression results for the prediction of post-harvest age (day 0-7 inclusive) in undamaged and damaged mushrooms

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Tissue</th>
<th>#Samples</th>
<th>#Loadings</th>
<th>RMSECV*</th>
<th>RER**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undamaged</td>
<td>Caps</td>
<td>80</td>
<td>7</td>
<td>1.2</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>Gills</td>
<td>80</td>
<td>7</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>Stalks</td>
<td>80</td>
<td>7</td>
<td>1.2</td>
<td>1.9</td>
</tr>
<tr>
<td>Damaged</td>
<td>Caps</td>
<td>80</td>
<td>7</td>
<td>1.3</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td>Gills</td>
<td>80</td>
<td>8</td>
<td>0.8</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>Stalks</td>
<td>80</td>
<td>6</td>
<td>1.2</td>
<td>2.2</td>
</tr>
</tbody>
</table>

*RMSECV= root mean square error of cross-validation (mean of 100 runs); **RER = SD/RMSECV

FIGURES

Figure 1 FTIR transmittance spectra of all mushroom tissues in (a) 400-1800 cm⁻¹, (b) 2800-3050 cm⁻¹, and (c) 3050–4000 cm⁻¹ wavenumber ranges
Figure 2. Average undamaged caps spectrum (raw data)
Figure 3 PC1 vs PC2 score plots of undamaged mushroom tissue (a) caps; (b) gills (c) stalks and damaged tissue (d) caps; (e) gills and (f) stalks: Sample ages from zero to seven
Figure 4(a) Relative importance plot of variables that are important in the random forest model for predicting damage/undamaged samples. The variable age being the most important followed by the wavenumbers 1868, 1870 and 1845 cm⁻¹. 4(b) Relative importance plot of variables that are important in the random forest model for predicting damaged/undamaged samples when age is not a variable. The most important variables are tissue type followed by the wavenumbers 1868, 1870 and 1560 cm⁻¹.
Figure 5. Frequency of generation of PLS regression models for mushroom post-harvest age on the basis of the number of latent variables selected. (a) undamaged caps, (b) undamaged gills, (c) undamaged stalks, (d) damaged caps, (e) damaged gills and (f) damaged stalks. Abscissa – no. of latent variables in model; ordinate – number of occurrences