Rapid Quality Assessment of Agricultural Products Using Spectroscopic Techniques

Zheng Bowen

Hunan Institute of Technology, Hengyang 421002, Hunan, China

Abstract: Spectroscopic techniques offer several advantages, including rapid detection, minimal sample loss, and the simultaneous analysis of multiple indicators. They can be applied to assess the sugar content, moisture content, protein content, maturity, mechanical damage, and varietal authenticity of agricultural products. This paper analyzes the mechanisms underlying the spectral responses of agricultural-product quality attributes and reviews methods for spectral acquisition, data preprocessing, feature selection, and model evaluation. It also discusses practical application conditions with reference to detection results obtained for grapes, winter jujubes, and animal-derived meat products. To address the limited applicability of models across varieties, substantial environmental interference, and insufficient stability in online detection, optimization measures are proposed, including expanding sample coverage, integrating multisource information, and improving instrument calibration.

Keywords: spectroscopic techniques; agricultural products; quality assessment; nondestructive testing; characteristic wavelengths

Introduction

Quality assessment of agricultural products is essential to production grading, storage management, processing control, and market supervision. Conventional physicochemical analyses generally require sampling, grinding, extraction, or titration. These procedures are time-consuming, and processed samples can rarely re-enter commercial circulation. Visual inspection can identify color, shape, and obvious physical damage, but it cannot accurately determine sugar content, moisture content, protein content, or internal defects. Bulk sorting operations also require detection systems to perform sampling, analysis, and classification continuously within a short period. Laboratory-based methods alone are therefore unable to satisfy the throughput requirements of commercial production.

Spectroscopic techniques record the absorption, reflection, transmission, or scattering of light at different wavelengths after it interacts with agricultural products, thereby establishing relationships between spectral characteristics and quality indicators. Visible spectroscopy is sensitive to variations in color and pigments, while near-infrared (NIR) spectroscopy can characterize the overtone and combination-band absorption of hydrogen-containing functional groups. Hyperspectral imaging simultaneously retains spectral and spatial information. Detection results are not derived directly from a single absorption peak; instead, predictions are generated through the combined use of reference measurements, spectral data, and mathematical models. Examining the technical workflow and application limitations of these techniques is therefore practically significant for improving the reliability of rapid quality assessment of agricultural products.

1. Technical Foundations of Spectroscopic Detection

1.1 Spectral Responses of Agricultural-Product Quality Attributes

Water, carbohydrates, proteins, fats, and organic acids in agricultural products contain different chemical bonds whose molecules produce absorption responses within specific wavelength regions. The NIR region principally records overtone and combination-band information generated by vibrations of C–H, O–H, and N–H bonds. Variations in moisture content alter the absorption intensity of bands associated with O–H bonds, while changes in carbohydrate and organic-acid concentrations produce subtle variations across multiple wavelength bands. The visible region mainly reflects the absorption characteristics of pigments such as chlorophylls, carotenoids, and anthocyanins and can therefore be used to evaluate maturity, surface color, and lesions.

The structural properties of agricultural tissues also affect light propagation. Pulp-cell size, intercellular spaces, peel thickness, and the extent of damage can all alter scattering paths. When mechanical damage occurs, localized tissue rupture, moisture migration, and changes in cellular structure produce differences in reflectance spectra. Spectra of agricultural products thus contain information on both chemical composition and physical structure, providing a basis for the integrated detection of internal components and surface defects.

1.2 Common Spectral-Acquisition Geometries

Diffuse reflectance is suitable for fruits, vegetables, grains, and powdered samples. When light illuminates a sample, part of the incident radiation is reflected from its surface, while another part penetrates the tissue and returns to the detector after absorption and scattering. Transmittance measurements can obtain information from deeper tissues and are suitable for assessing the internal quality of fruit. However, signal intensity is affected by light-source power, sample size, and peel thickness. In interactance measurements, the light source and detector are positioned adjacent to each other. This arrangement reduces direct surface reflection while maintaining a certain penetration depth.

Hyperspectral imaging acquires images over a sequence of contiguous narrow wavelength bands, with each pixel corresponding to a complete spectrum. This technique can both identify spectral differences in abnormal regions and determine the spatial locations of defects. It is therefore suitable for detecting bruising, fungal decay, insect infestation, and foreign materials. However, hyperspectral imaging generates large volumes of data and imposes substantial computational demands. Production-line applications generally require the selection of a limited number of characteristic wavelengths, followed by the development of a multispectral detection system.

2. Principal Detection Targets of Spectroscopic Techniques

2.1 Detection of Internal Physicochemical Quality

Sugar content, moisture content, protein content, fat content, acidity, and starch content are among the principal indicators used in quantitative spectroscopic analysis. During model development, physicochemical reference values are first obtained using standard methods, after which the statistical relationship between the spectral matrix and the reference values is analyzed. In a study on the determination of soluble solids content (SSC) in grapes, a model was developed using the wavelength range of 408–1092.8 nm. The coefficient of determination for the combined prediction set reached 0.9808, and the standard error of prediction was 0.354 °Brix. The model covered three grape varieties: Muscat Hamburg, Manai, and Red Globe [2]. These results demonstrate that visible–NIR spectroscopy can predict sugar content across multiple fruit varieties when an appropriate sample range and wavelength combination are employed.

Grains and powdered agricultural products are relatively homogeneous in form, and their spectra generally exhibit better repeatability than those of intact fruit. Spectroscopic techniques can therefore be used to determine their moisture, protein, fat, and starch contents. By contrast, fruit and vegetable surfaces exhibit more complex curvature, peel color, and internal tissue distributions. The measurement position should consequently be standardized, and localized differences should be reduced by rotating the sample or conducting measurements at multiple points.

2.2 Identification of External Defects and Mechanical Damage

Conventional imaging primarily records visible color and texture and has limited ability to identify early-stage damage before discoloration occurs. Hyperspectral imaging can distinguish damaged areas on the basis of changes in moisture status, tissue structure, and light scattering. Sun et al. acquired hyperspectral images of winter jujubes over the range of 900–1700 nm and selected characteristic wavelengths near 1353 and 1691 nm. A support vector machine achieved an identification accuracy of 95.16% for damaged regions [3].

Fungal decay, insect infestation, and epidermal lesions also alter spectral reflectance in specific regions. A detection system can first segment the sample from the background, subsequently extract abnormal pixels, and ultimately generate a defect-distribution map. Compared with point spectroscopy, hyperspectral imaging offers clear advantages in defect localization. Nevertheless, its processing speed and equipment cost remain important constraints on online applications.

2.3 Identification of Variety, Geographical Origin, and Adulteration

Differences in variety, geographical origin, and rearing conditions can produce variations in the compositional ratios and tissue structures of agricultural products. Such differences may not correspond to a distinct individual absorption peak but may generate stable combinations of characteristics across the full spectrum. Principal component analysis, discriminant analysis, and support vector machines can extract differences among sample categories and thereby support varietal identification, geographical-origin traceability, and adulteration screening.

Meng et al. used Fourier-transform near-infrared spectroscopy to distinguish pork, beef, and mutton, acquiring spectra over the range of 9881.46–4119.20 cm^{-1}. In the prediction set, the identification accuracies for the three meat categories were 99.28%, 97.42%, and 100%, respectively [4]. These results demonstrate that spectroscopic techniques can be used to identify categories of agricultural products. However, high accuracy under specific sample, instrument, and experimental conditions cannot be directly interpreted as equivalent performance across regions and production batches.

Table 1. Representative Results of Spectroscopic Detection of Agricultural Products

Detection target	Spectral range or characteristic wavelengths	Detection task	Principal result
Grapes	408–1092.8 nm	Prediction of soluble solids content	Prediction coefficient of determination of 0.9808 and standard error of prediction of 0.354 °Brix
Winter jujubes	1353 and 1691 nm	Identification of early mechanical damage	Support vector machine identification accuracy of 95.16%
Pork, beef, and mutton	9881.46–4119.20 cm^{-1}	Identification of animal-source categories	Prediction accuracies of 99.28%, 97.42%, and 100%, respectively

3. Development of Models for Agricultural-Product Quality Assessment

3.1 Sample Design and Determination of Reference Values

Sample representativeness determines the scope of model applicability. Samples should cover different varieties, production regions, maturity levels, storage periods, and quality gradients. Quantitative detection additionally requires a sufficiently broad distribution of indicator values. If samples are concentrated within a narrow range, a model may lack practical predictive ability even when it achieves a high coefficient of determination. The calibration and prediction sets should be independent of each other while maintaining comparable quality distributions.

After spectral acquisition, reference values should be determined using national standard methods, generally accepted industry methods, or validated instruments. Sugar content may be determined through refractometry, moisture content through direct oven drying, and protein content through the Kjeldahl method. Errors in reference values are propagated into the resulting model. Replicate measurements, instrument calibration, and the review of anomalous data are therefore indispensable.

3.2 Spectral Preprocessing and Feature Selection

Raw spectra are readily affected by baseline drift, particle scattering, surface curvature, and instrument noise. Smoothing can reduce random noise, while first- or second-order derivatives can separate overlapping bands and correct baseline variations. Standard normal variate transformation and multiplicative scatter correction are frequently used to reduce scattering differences. Excessive preprocessing may, however, amplify noise or remove useful information. Appropriate preprocessing combinations should therefore be selected through cross-validation.

The successive projections algorithm, competitive adaptive reweighted sampling, and variable importance in projection can be used to extract informative variables from full-spectrum data. Reducing the number of wavelengths can lower model complexity and provide a basis for designing low-cost multispectral instruments. Feature selection must be performed exclusively within the training data. If all samples are used in advance to select wavelengths, information leakage may occur and lead to overly optimistic prediction results.

3.3 Model Development and Performance Evaluation

Partial least-squares regression is suitable for quantitative problems involving large numbers of spectral variables and strong correlations among those variables. Support vector machines can model certain nonlinear relationships and are appropriate for classification and regression tasks involving limited sample sizes. Random forests possess a strong capacity to combine features, while neural networks can learn complex relationships. However, neural networks require large and representative datasets and present difficulties in parameter tuning and result interpretation.

Quantitative models can be evaluated using the coefficient of determination, root-mean-square error, standard error of prediction, and ratio of performance to deviation. Classification models should report accuracy, sensitivity, specificity, and a confusion matrix. Results obtained solely from the training set cannot demonstrate model generalizability. Independent prediction sets, cross-batch validation, and cross-season validation provide more meaningful evidence of practical applicability.

4. Principal Challenges in Practical Applications

4.1 Effects of Sample Variability on Model Stability

Agricultural products exhibit substantial biological variability. Differences in geographical origin, climate, maturity, and storage conditions within the same variety can cause shifts in spectral distributions. A laboratory model containing samples from only one production region or harvest season may produce substantial prediction errors when applied to a new batch. Fruit temperature, surface moisture, orientation, and measurement position can also alter spectral signals.

Model updating should not simply involve increasing the number of samples. It should also address quality ranges that were insufficiently represented in the original model. Anomalous batches can be identified using spectral distance or prediction residuals, and representative new samples can then be incorporated into the calibration set to support periodic model revision.

4.2 Limitations in Transferring Laboratory Models to Online Detection

Laboratory acquisition is generally conducted using a stable light source, a fixed measurement distance, and relatively long exposure times. Production lines, by contrast, are affected by conveyor vibration, variations in ambient light, and random sample orientation. When a model that performs well under laboratory conditions is transferred to online equipment, changes in spectral resolution, detector response, and optical-path configuration may reduce its predictive accuracy.

The large data volume generated by hyperspectral equipment is unfavorable for high-speed sorting. In production applications, filters, light-emitting diodes, or compact spectrometer modules can be configured according to selected characteristic wavelengths to reduce redundant data. Standardized calibration-transfer relationships must also be established among instruments to prevent the same model from producing systematic deviations when applied to different devices.

5. Directions for Optimizing Rapid Spectroscopic Detection

5.1 Expanding Sample Coverage

Modeling samples should include multiple production regions, varieties, harvest seasons, and storage stages, and complete records of measurement conditions should be maintained. Samples may be divided according to production region or batch to prevent specimens from the same batch from appearing simultaneously in the calibration and prediction sets. External validation across different years can be used to evaluate a model's resistance to seasonal variation. Research on the NIR-based nondestructive testing of citrus fruit has likewise identified instrument parameters, chemometric methods, and model applicability as important factors affecting detection performance [1].

5.2 Promoting Multisource Information Fusion

Spectral data are suitable for analyzing chemical composition, whereas conventional images are effective for recording color, shape, and texture. Combining these two types of information can facilitate the simultaneous assessment of external grade and internal quality. Data from temperature, weight, firmness, and odor sensors can also serve as auxiliary variables, thereby improving the identification of complex quality conditions. Information-fusion approaches should control the number of variables and analyze the actual contribution of each data type to prediction performance to avoid overfitting as model size increases.

5.3 Improving Portable and Online Detection Systems

Portable instruments are suitable for field-based maturity assessment, quality inspection during procurement, and storage monitoring, whereas online equipment is intended for processing and grading operations. System design must balance detection accuracy, acquisition time per sample, equipment cost, and maintenance requirements. Models should specify their applicable ranges and include warnings for anomalous samples. When a sample spectrum falls outside the calibration range, the system should refrain from directly reporting a quality value and instead recommend confirmatory testing. This approach can reduce the influence of erroneous results on production decisions.

Conclusion

Spectroscopic techniques use differences in the absorption, reflection, and scattering of agricultural products across different wavelength bands to rapidly assess internal composition, maturity, mechanical damage, and product authenticity. Studies involving grape sugar content, mechanical damage in winter jujubes, and the identification of animal-derived meat products have achieved high experimental accuracy, demonstrating the potential of spectroscopic analysis for nondestructive grading and quality control. Detection performance nevertheless depends on sample representativeness, the quality of reference measurements, spectral preprocessing, and model-validation procedures.

Future studies should strengthen validation across production regions, seasons, and instruments; promote the integrated analysis of spectral, imaging, and other sensor data; and select characteristic wavelengths according to production-line throughput requirements. Only after stable mechanisms for instrument calibration, model updating, and confirmatory testing of anomalous samples have been established can spectroscopic detection progress from laboratory analysis to a reliable production tool.

References

[1] Zhang, X. X., Li, P., Yu, M., et al. Advances in near-infrared spectroscopic techniques for the nondestructive testing of citrus fruit [J]. Food Science, 2022, 43(1), 260–268.

[2] Chen, C., Lu, X. X., Zhang, P., et al. Quantitative prediction of soluble solids content in grapes during storage using visible–near-infrared diffuse reflectance spectroscopy [J]. Food Science, 2015, 36(20), 109–114.

[3] Sun, S. P., Peng, J., Li, R., et al. Early detection of mechanical damage in winter jujubes using near-infrared hyperspectral imaging [J]. Food Science, 2017, 38(2), 301–305.

[4] Meng, Y., Zhang, Y. H., Wang, J. M., et al. Rapid identification of meat products from different animal sources using near-infrared spectroscopy [J]. Food Science, 2014, 35(6), 156–158.