Review

Research on Nondestructive Inspection of Fruits Based on Spectroscopy Techniques: Experimental Scenarios, ROI, Number of Samples, and Number of Features

Qi Wang 1,2, Jinzhu Lu 1,2,*, Yuanhong Wang 1,2 and Junfeng Gao 3

1 Modern Agricultural Equipment Research Institute, School of Mechanical Engineering, Xihua University, Chengdu 610039, China; 1202208550098@stu.xhu.edu.cn (Q.W.); wangyuanhong@stu.xhu.edu.cn (Y.W.)
2 School of Mechanical Engineering, Xihua University, Chengdu 610039, China
3 Department of Computer Science, University of Aberdeen, Aberdeen AB24 3FX, UK; junfeng.gao@abdn.ac.uk
* Correspondence: lujinzhu@mail.xhu.edu.cn

Abstract: Spectral technology is a scientific method used to study and analyze substances. In recent years, the role of spectral technology in the non-destructive testing (NDT) of fruits has become increasingly important, and it is expected that its application in the NDT of fruits will be promoted in the coming years. However, there are still challenges in terms of dataset collection methods. This article aims to enhance the effectiveness of spectral technology in NDT of citrus and other fruits and to apply this technology in orchard environments. Firstly, the principles of spectral imaging systems and chemometric methods in spectral analysis are summarized. In addition, while collecting fruit samples, selecting an experimental environment is crucial for the study of maturity classification and pest detection. Subsequently, this article elaborates on the methods for selecting regions of interest (ROIs) for fruits in this field, considering both quantitative and qualitative perspectives. Finally, the impact of sample size and feature size selection on the experimental process is discussed, and the advantages and limitations of the current research are analyzed. Therefore, future research should focus on addressing the challenges of spectroscopy techniques in the non-destructive inspection of citrus and other fruits to improve the accuracy and stability of the inspection process. At the same time, achieving the collection of spectral data of citrus samples in orchard environments, efficiently selecting regions of interest, scientifically selecting sample and feature quantities, and optimizing the entire dataset collection process are critical future research directions. Such efforts will help to improve the application efficiency of spectral technology in the fruit industry and provide broad opportunities for further research.

Keywords: fruits; spectroscopy techniques; non-destructive testing; experimental environment; ROI; sample size; number of spectral features

1. Introduction

Fruits are an essential part of our daily life and are rich in nutrients. For example, citrus, which mainly grows in hot and subtropical regions, has become one of the main fruits in the world because it is rich in vitamin C, vitamin B6, dietary fiber, sugar, potassium, calcium, and other advantages [1,2]. With the continuous improvement of China’s national living standards, China’s fruit production has shown an apparent upward trend. According to the data released by the National Bureau of Statistics, it grew from 256,883,500 tons in 2018 to 327,442,800 tons in 2023, with sufficient growth momentum [3].
Fruits are becoming increasingly important in our daily lives, and the research on non-destructive testing (NDT) of fruits is also receiving more attention. The importance of NDT for fruits cannot be ignored. Firstly, NDT helps to ensure the safety and health of fruits. For example, through testing, harmful residues or pests in fruits can be detected promptly, and remedial measures can be taken to protect consumers’ health rights and interests and enhance consumers’ trust in fruits [4]. Secondly, NDT also helps to improve the commercialization level of fruits. For example, by testing fruits’ size, weight, and internal component content, it is possible to screen out fruits with beautiful appearance and good quality and enhance their added value in the market [5]. In addition, non-destructive testing of fruits is crucial for their storage periods [6]. Regular quality testing can ensure that the quality of fruits is effectively maintained during storage. For example, quality testing can promptly identify critical indicators in fruits, such as ripeness, texture, and sugar content, which directly affect the storage period of fruits. Based on the test results, corresponding storage measures can be taken, such as adjusting the storage temperature, humidity, ventilation, and other conditions to extend the storage periods of fruits [7]. Most importantly, NDT is an essential means to ensure the sustainable development of the fruit industry. Only by ensuring that the quality of fruits is stable can we maintain the sustained consumer demand and promote the citrus industry’s healthy development. Therefore, strengthening quality testing and improving fruit quality plays an irreplaceable role in guaranteeing the sustainable development of the fruit industry.

Spectroscopy is a scientific method used to study and analyze states of matter involving light interactions and dispersion. This technique utilizes the properties of substances in terms of absorption, scattering, or emission of light at different wavelengths to obtain information about the composition, structure, and properties of substances. Spectroscopy techniques can be categorized into various types. Some of the common ones include hyperspectral, near-infrared spectroscopy, mid-infrared spectroscopy, Raman spectroscopy, ultraviolet-visible (UV-Vis) spectroscopy, molecular fluorescence spectroscopy, etc. [8]. These spectroscopy techniques have a wide range of applications in fields such as agronomy, medicine, physics, and environmental sciences, providing researchers with tools to gain insight into the properties of substances.

In quality testing, spectroscopy techniques can obtain spectral profiles by measuring light reflected or absorbed at specific wavelengths by agricultural products. These profiles contain spectral characteristics of the substance of interest, such as the soluble solid content (SSC) [9], sugar content [10], and water content [11] of the fruit. By analyzing these features, product quality can be assessed quickly and accurately. Produce with different maturity levels exhibit different characteristics on the spectrum. Spectroscopy techniques can determine the maturity of fruits, vegetables, or other agricultural products to select the appropriate harvest time and improve the quality of the products [12,13]. In pesticide residue detection, spectroscopic technology can identify the unique fingerprints of different pesticides on the spectrum. By collecting spectral data from a sample and comparing it to a library of spectra of known pesticides, the presence of pesticide residues in agricultural products can be detected. Regarding nutrient analysis, spectral technology can detect the nutrients of farm products, such as proteins, fats, and sugars. Since different components exhibit different spectral features at different wavelengths, spectroscopic techniques are analyzed by these features to provide accurate component contents. For example, Choi J.-H. et al. collected spectral data using NIR spectroscopic techniques followed by regression analysis of the sugar content of pears using a partial least squares regression model, which showed that the calibrated relative standard error or predicted relative standard error (rSEC/rSEP) = 0.91–0.92 [14]. The nutrient content of agricultural products directly affects their quality. By analyzing the nutrient content through spectroscopic technology, farmers and producers can implement fine quality control to ensure that the products meet market demands and standards. Regarding disease detection in agricultural products, their tissues show abnormal spectral characteristics when plants are infected.
with diseases. Spectroscopy can detect these abnormalities and help in the early detection of plant diseases. Spectroscopy techniques can provide information about the type and extent of plant diseases. This helps farmers to take timely measures to reduce the spread of diseases and protect the growth and yield of agricultural products [15]. Overall, applying spectroscopy techniques in these areas provides agriculture with efficient, non-destructive tools that help farmers and producers better manage agricultural products’ production, quality, and safety.

The application of spectroscopic technology in fruit production can effectively improve aspects of production efficiency such as ripeness classification, detection of physicochemical properties, defect detection, pest and disease detection, and other aspects. Many aspects of fruit production can be further developed under the impetus of this technology. This paper reviews the selection of the environment, the selection region of interest, the selection of the number of samples, and the number of spectral features of the experiments of spectroscopic techniques in the process of non-destructive testing of citrus and other fruits. It also analyzes the current problems and prospects for future research to achieve the collection of spectral data of fruit samples in orchard environments, the more efficient selection of the region of interest, and the scientific selection of the number of samples and the number of features.

2. Spectral Imaging Technique

2.1. Principles of Spectroscopy

Spectral imaging technology combines imaging and spectral measurement techniques, simultaneously acquiring image and spectral information to form a “data cube” containing spectral radiation information in two dimensions and wavelength distribution. Figure 1 shows a schematic of a hyperspectral reflectance imaging system based on a laboratory environment. Since traditional single-band photodetection spectrometers are only capable of acquiring specified wavelength or bandwidth photon signals, they provide very little information. In contrast, hyperspectral techniques can provide richer information about the target scene [16].

![Schematic diagram of reflectance hyperspectral imaging system.](image)

According to the number of spectral bands and spectral resolution, spectral imaging technology can be roughly divided into three categories: (1) multispectral imaging technology, where the acquired image data has only a few spectral bands, the spectral resolution is generally about 100 nm, and the multispectral imager is usually called a multispectral camera [17]; (2) hyperspectral imaging technology, with dozens of spectral bands, where the spectral resolution is generally about 10 nm [18]; (3) and ultra-spectral imaging
technology. Acquired image data usually have several hundred spectral bands, and the spectral resolution is usually below 1 nm [19].

Multi-spectral imaging technology acquires the spectral information of the same target in different bands synchronously by splitting the electromagnetic wave radiation into several narrow spectral bands and scanning it. Multi-spectral imaging technology shows its excellent application value when facing complex and changing external environments and fruits with different shapes. The technology can simultaneously process visible and infrared spectral images to capture the external and internal feature information of fruit, and to realize precise detection and diagnostic research on the growth conditions of fruits [20,21].

Among these three categories, hyperspectral imaging techniques in fruits has been extensively studied [22-24]. Hyperspectral imaging combines imaging and spectroscopic techniques to capture continuous spectral data of a specific object and generate spectral images. Imaging technology provides visual information, spectral technology reveals spectral properties, and the integrated approach accurately resolves the composition and chemical properties of substances, providing a basis for analyzing and evaluating target properties. Hyperspectral imaging provides richer spatial, visual, and spectral information than traditional techniques [25].

Ultra-spectral imaging is a technique that utilizes imaging spectrometers to acquire many very narrow, spectrally continuous image data in the visible (400–700 nm), near-infrared (700–2500 nm), mid-infrared (2500–100,000 nm) wavelength ranges of the electromagnetic spectrum [26]. This technique has a wide range of applications in remote sensing, medical imaging, agriculture, and environmental monitoring [27]. As it provides detailed spectral information, it helps to identify and analyze the target substance’s chemical composition and structural properties. For example, Antony, M.M. et al. [19] used spectral libraries generated by a hyperspectral imaging system to automatically classify healthy abnormal samples of cabbage.

2.2. Spectral Data Acquisition and Processing

In spectral imaging technology, relying solely on spectral and image information for qualitative analysis is insufficient. Although such an analysis can provide some basic information about the object, such as its color, shape, and specific physical properties, this information is usually limited, and it is difficult to reveal the specific characteristics and changing patterns of the object in depth [28]. Quantitative testing is necessary to gain a more comprehensive understanding of the intrinsic qualities of the object under study. Quantitative testing can provide more accurate and detailed data, allowing researchers to analyze the sugar level [29], SSC [30], acidity [31], etc., of an object. This not only helps to improve the accuracy of the analysis, but also provides a solid foundation for further research and applications. Therefore, for spectral and image data, we need to analyze them quantitatively and build corresponding mathematical models to make more accurate predictions about unknown research objects. Such a quantitative approach provides deeper insights and a more reliable scientific basis for quality assessment and research. Therefore, many studies on quantitative analysis of fruits have been carried out by domestic and foreign researchers [32].

In the process of acquiring spectra and images, to minimize the spatial intensity variation of light and the dark current effect of the camera, white light and dark light are used to correct the reflectance of hyperspectral images of all navel orange samples. The calibration formula is shown in Equation (1).
where $R$ denotes the relative reflectance value at wavelength $\lambda$, and $n$ is a pixel. $S$ and $B$ represent the sample image and white reference image, respectively, and $D$ is the dark image [33].

After collecting the spectral information of the samples, the region of interest was selected according to the demand using threshold segmentation, point selection, box selection, or global selection, and the reflectance spectral curves in the region of interest can be calculated. Finally, the compositional content information of the samples was collected by a digital refractometer, PH meter, and hardness meter in a laboratory environment, and the common component contents were sugar, acidity, hardness, vitamin C, etc. [34,35]. Figure 2 shows the spectral and component information acquisition steps for the samples.

$$R_{\lambda,n} = \frac{S_{\lambda,n} - D_{\lambda,n}}{B_{\lambda,n} - D_{\lambda,n}}$$ (1)

![Diagram](https://example.com/diagram.png)

**Figure 2.** Spectral and component information acquisition steps for samples.
Spectroscopy is an important aspect of chemical analysis. Qualitative and quantitative analysis of spectroscopic data using chemometric methods is necessary in most cases, and the discriminant or regression models established by chemometric methods are the basis for industrial applications. Spectra reflect information about the interatomic vibrations of functional groups. By analyzing the characteristic spectral lines, it is possible to determine the functional groups contained in the object of study and the corresponding molecular or material information [36]. In the field of spectral analysis, chemometric methods are widely used. As an important aspect of chemical analysis, the qualitative and quantitative analysis of spectral data by chemometric methods is necessary in most cases, and the discriminant or regression models established by chemometric methods are the basis for industrial applications. Spectra reflect information about the interatomic vibrations of functional groups. By analyzing the characteristic spectral lines, it is possible to determine the functional groups contained in the object of study and the corresponding molecular or material information. It is an indirect means of analysis that requires a calibration model with known samples to predict the composition or concentration of unknown samples. Multivariate correction is a statistical technique for modeling relationships between multiple variables, often used when analyzing and interpreting complex data sets. When performing multivariate correction, we rely on information about the components of known samples to construct the correction model [37]. In the model identification phase, we use supervised or unsupervised discriminant analysis methods to build identification models based on spectral data. Supervised discriminant analysis relies on labeled data to train the model to distinguish between different classes. In contrast, unsupervised discriminant analysis does not rely on labeled data, but instead classifies and identifies by discovering the intrinsic structure of the data [38]. For unknown samples, whether it is to determine their composition or concentration or to determine their category, it is necessary to apply the established multivariate correction or recognition model combined with the spectral data of the unknown samples for prediction and recognition.

To construct robust, accurate, reliable, and representative correction or identification models, the spectrometer's performance is highly required, the samples selected should be fully representative, and the influence of spectral acquisition conditions needs to be considered. Figure 3 depicts the operational flow of chemometric methods in spectral analysis, and these steps are critical for constructing high-quality and robust analytical models. Specifically, determining spectral range requires selecting an appropriate wavelength range based on the properties of the analyzed substance to ensure the capture of critical information [39]. Spectral preprocessing strategies, such as smoothing, denoising, baseline correction, and standard normal transform (SNV), help to improve spectral data quality and consistency [40]. Outlier handling can ensure model robustness through outlier detection and removal techniques. Multivariate correction methods, such as principal component regression (PCR) and partial least squares regression (PLSR), help to establish accurate relationships between independent and dependent variables [41]. Pattern recognition methods, both supervised and unsupervised, are used for sample classification and pattern recognition. Feature wavelength selection through stepwise regression and genetic algorithms (GA) selects the most representative wavelengths, simplifying the model and improving its performance. Spectral data compression, e.g., through principal component analysis (PCA), reduces data dimensionality and computational complexity [42]. Each step, from spectral acquisition to data preprocessing to model building and validation, is critical to ensure the accuracy and reliability of the final model.
3. Experimental Environment

The application of spectroscopic technology in fruit inspection is an important research direction. It can be used to analyze the spectral characteristics of fruits to achieve rapid and accurate detection of fruit quality, ripeness, and nutrient content. Research in this area usually involves both laboratory environments and natural orchard environments. For example, Figure 4a shows the spectral images of peach samples obtained by Li J. et al. [43] using a hyperspectral instrument in a laboratory environment, and Figure 4b shows visible-near-infrared hyperspectral images of apples measured by Wang F. et al. [44] using a hyperspectral imager in an orchard environment.

Figure 3. Operational procedures for chemometric methods.

Figure 4. Experimental environments: (a) laboratory environment; (b) orchard environment.
3.1. Laboratory Environment

Researchers have made great strides in the study of citrus in laboratory settings, as shown in Table 1.

**Table 1. Application of spectroscopic techniques for nondestructive testing of citrus under laboratory conditions.**

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Fruit (No. of Samples)</th>
<th>Quality Parameters</th>
<th>Wavelength Range (nm)</th>
<th>Light Source</th>
<th>Data Analysis</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>[45]</td>
<td>Navel oranges (126 samples)</td>
<td>Defects</td>
<td>975.15–2196.2</td>
<td>Two 500 W halogen lamps</td>
<td>QCF</td>
<td>Accuracy = 100%</td>
</tr>
<tr>
<td>[46]</td>
<td>Oranges (460 samples)</td>
<td>Defects</td>
<td>400–1000</td>
<td>Two 150 W halogen lamps</td>
<td>BR + PCA</td>
<td>Accuracy = 93.7%</td>
</tr>
<tr>
<td>[47]</td>
<td>Citrus (606 samples)</td>
<td>Defects</td>
<td>400–1000</td>
<td>Two 100 W tungsten halogen lamps</td>
<td>PCA</td>
<td>Accuracy = 96.63%</td>
</tr>
<tr>
<td>[48]</td>
<td>Citrus (426 samples)</td>
<td>Mold infection Detection of fungal infections</td>
<td>900–1700</td>
<td>20 W halogen lamp</td>
<td>PLS-DA</td>
<td>Accuracy = 100%</td>
</tr>
<tr>
<td>[49]</td>
<td>Citrus (117 samples)</td>
<td>SSC</td>
<td>1000–1700</td>
<td>10 W halogen light source</td>
<td>FA</td>
<td>Accuracy = 97.8%</td>
</tr>
<tr>
<td>[50]</td>
<td>Nanfeng mandarin (160 samples)</td>
<td>SSC</td>
<td>380–1030</td>
<td>Two 150 W halogen lamps</td>
<td>PLSR</td>
<td>$R_p^2 = 0.956$</td>
</tr>
<tr>
<td>[51]</td>
<td>Citrus (150 samples)</td>
<td>SSC</td>
<td>200–1100</td>
<td>-</td>
<td>MLR</td>
<td>$R_p = 0.893$</td>
</tr>
<tr>
<td>[52]</td>
<td>Navel orange (150 samples)</td>
<td>Maturity</td>
<td>390–1055</td>
<td>Four 50 W halogen lamps</td>
<td>KNN</td>
<td>Accuracy = 96%</td>
</tr>
<tr>
<td>[53]</td>
<td>Citrus (105 samples)</td>
<td>SSC</td>
<td>600–1100</td>
<td>Light source of 12 V/20 W</td>
<td>PLS</td>
<td>$R_p = 0.82$</td>
</tr>
</tbody>
</table>


We observed the findings. In terms of qualitative analysis, citrus information can be predicted very well, especially regarding defect detection, and up to 100% detection accuracy was achieved. In terms of quantitative analysis, the prediction of the SSC of citrus also showed excellent results.

In addition to citrus fruits, other fruits, such as peaches, bananas, apples, kiwis, etc., have also been studied in terms of non-destructive testing based on spectroscopy techniques under laboratory conditions. All of them have achieved excellent predictive results. As shown in Table 2, researchers from various countries have also conducted studies on defect detection, quantitative analysis, and qualitative analysis of other fruits under laboratory conditions using spectroscopy techniques. Among the indices studied, we found that in terms of quantitative analysis, more studies were conducted on the SSC, brix, hardness, and acidity indices. In light source selection, as with citrus testing, most of the light sources used in their experiments were halogen lamps, only perhaps with slight differences in quantity and wattage. From the results of the study, the performances of their evaluation indices were excellent under laboratory conditions, and it is evident that spectroscopy techniques have matured in this environment for fruit quality detection.
Table 2. Application of spectroscopy techniques to non-destructive testing of fruit under laboratory conditions.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Fruit (No. of Samples)</th>
<th>Experimental Environment</th>
<th>Quality Parameters</th>
<th>Wavelength Range (nm)</th>
<th>Light Source</th>
<th>Data Analysis</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>[54]</td>
<td>Pear (279 samples)</td>
<td>Laboratory</td>
<td>SSC</td>
<td>397–1187</td>
<td>150 W halogen tungsten lamp</td>
<td>MLP-CNN-TCN</td>
<td>$R_p^2 = 0.956$</td>
</tr>
<tr>
<td>[55]</td>
<td>Grape (360 samples)</td>
<td>Laboratory</td>
<td>SSC</td>
<td>450–1000</td>
<td>Four 50 W halogen lamps</td>
<td>PLSR</td>
<td>$R_p = 0.9762$</td>
</tr>
<tr>
<td>[56]</td>
<td>Kiwifruit (100 samples)</td>
<td>Laboratory</td>
<td>SSC, PH</td>
<td>400–1000</td>
<td>A tungsten halogen lamp 300 W</td>
<td>PLSR</td>
<td>$R_p = 0.93,0.943$</td>
</tr>
<tr>
<td>[57]</td>
<td>Honey peach (300 samples)</td>
<td>Laboratory</td>
<td>Chlorophyll content</td>
<td>400–1000</td>
<td>Regulated halogen tungsten lamp (0–150 w)</td>
<td>PLS</td>
<td>$R_p = 0.904$</td>
</tr>
<tr>
<td>[58]</td>
<td>Peach (500 samples)</td>
<td>Laboratory</td>
<td>Five peach varieties</td>
<td>350–820</td>
<td>Tungsten halogen lamp (3.5 w)</td>
<td>CNN</td>
<td>Accuracy = 100%</td>
</tr>
<tr>
<td>[59]</td>
<td>Peach (200 samples)</td>
<td>Laboratory</td>
<td>Two peach varieties</td>
<td>400–2500</td>
<td>Tungsten halogen lamp(400 W)</td>
<td>GSR</td>
<td>Accuracy = 99.3%</td>
</tr>
<tr>
<td>[60]</td>
<td>Apple (60 samples)</td>
<td>Laboratory</td>
<td>Bruises</td>
<td>381–1037</td>
<td>Four 75 W tungsten halogen lamps</td>
<td>CNN</td>
<td>Accuracy = 95.79%</td>
</tr>
<tr>
<td>[61]</td>
<td>mango</td>
<td>Laboratory</td>
<td>Mechanically induced damage</td>
<td>650–1100</td>
<td>Twelve halogen lamps (20 W)</td>
<td>KNN</td>
<td>Accuracy = 97.9%</td>
</tr>
<tr>
<td>[62]</td>
<td>Mango (161 samples)</td>
<td>Laboratory</td>
<td>Anthracnose</td>
<td>350 –1900</td>
<td>Two halogen bulbs (120 V 400 W)</td>
<td>LDA</td>
<td>Accuracy = 91–100%</td>
</tr>
<tr>
<td>[63]</td>
<td>Mango (40 samples)</td>
<td>Laboratory</td>
<td>Disease</td>
<td>1000–2500</td>
<td>-</td>
<td>PCA</td>
<td>Accuracy = 99%</td>
</tr>
<tr>
<td>[64]</td>
<td>Banana (330 samples)</td>
<td>Laboratory</td>
<td>Colletotrichum species</td>
<td>339–1019</td>
<td>-</td>
<td>CNN</td>
<td>Accuracy = 97.37%</td>
</tr>
<tr>
<td>[65]</td>
<td>Watermelon (200 samples)</td>
<td>Laboratory</td>
<td>Maturity</td>
<td>200–1110</td>
<td>Tungsten halogen lamps</td>
<td>Corrected-RPP (C-RPP)</td>
<td>CRR = 88.1%</td>
</tr>
</tbody>
</table>


In summary, the use of spectroscopy techniques in a laboratory setting has demonstrated effectiveness in detecting defects in citrus and other fruits, as well as in quantitative analysis and ripeness classification. Researchers can establish highly controlled conditions to ensure the stability and reproducibility of experiments. This environment enables researchers to utilize high-precision spectroscopic instruments for the comprehensive spectroscopic analysis of fruits, yielding accurate data. Moreover, the laboratory can prepare various standard samples for calibrating instruments, verifying model reliability, and conducting model building and optimization.
3.2. Orchard Environment

Laboratory environments are typically artificially controlled and differ from real orchard growing environments. Such differences may lead to significant deviations between the results obtained in the laboratory and the actual application scenarios [66]. In addition, due to the complexity and variability of field environments, laboratory environments often do not fully simulate external influences in the orchards, such as changes in sunlight, wind speed, and temperature. These factors can have an impact on the acquisition of fruit spectra. For example, in the detection of citrus defects, in addition to some post-picking falls, friction injuries, freezing injuries, etc., the samples collected in the laboratory environment do not take into account the effects of pre-picking diseases, insect pests, poor growth, and weather factors. They, therefore, may have significant errors when validated in the orchard environment [67]. In citrus ripeness detection, researchers typically select samples of different ripeness levels for experiments, and the spectral information of the samples can be collected more accurately in the laboratory. However, in the actual application process, to determine the maturity of the samples, the influencing factors in the orchard environment need to be considered, so the researchers need to measure the spectra of the samples in the orchard environment to satisfy the classification of citrus maturity [68]. Currently, recognizing that measuring certain fruits in the laboratory environment may not meet practical application requirements, some researchers have explored fruit quality testing in the orchard environment. Two types of fruits, berries and drupes, are discussed below.

3.2.1. Berries

Berries are a group of small, juicy, colorful, and nutrient-rich fruits that usually includes strawberries, grapes, and blackberries [69]. Shown in Table 3 are the researchers' studies based on spectroscopic techniques in a berry-growing environment.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Fruit (No. of Samples)</th>
<th>Quality Parameters</th>
<th>Wavelength Range (nm)</th>
<th>Light Source</th>
<th>Data Analysis</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>[70]</td>
<td>Grapes (429 samples)</td>
<td>SSC</td>
<td>400–1000</td>
<td>Clear sky conditions (10:30 a.m. to 12:00 p.m.)</td>
<td>PLS</td>
<td>(R^2 = 0.77)</td>
</tr>
<tr>
<td>[71]</td>
<td>Grapes (90 trees)</td>
<td>Sugar content</td>
<td>450–1000</td>
<td>Clear sky conditions (9:00 a.m. to 14:00 p.m.)</td>
<td>Novel autoencoder-based framework</td>
<td>(R^2 = 0.7)</td>
</tr>
<tr>
<td>[72]</td>
<td>Strawberry (120 samples)</td>
<td>Maturity</td>
<td>370–1015</td>
<td>Overcast sky</td>
<td>CNN</td>
<td>Accuracy = 98.6%</td>
</tr>
</tbody>
</table>


3.2.2. Drupes

Drupes are a group of fruits with a hard core in the center and a juicy outer layer, usually including apples, mangoes, and peaches [73]. As shown in Table 4, the researchers collected spectral data in a drupe-growing environment and based their studies on the spectroscopic techniques which were used.
Agriculture 2024, 14, 977 11 of 23

Table 4. Application of spectroscopic techniques for the non-destructive detection of drupes in an orchard environment.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Fruit (No. of Samples)</th>
<th>Quality Parameters</th>
<th>Wavelength Range (nm)</th>
<th>Light Source</th>
<th>Data Analysis</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>[44]</td>
<td>Apple (100 samples)</td>
<td>Firmness</td>
<td>500–900</td>
<td>Natural light</td>
<td>PLSR</td>
<td>$R_v = 0.783$</td>
</tr>
<tr>
<td>[44]</td>
<td>Apple (100 samples)</td>
<td>SSC</td>
<td>500–900</td>
<td>Natural light</td>
<td>PLSR</td>
<td>$R_v = 0.901$</td>
</tr>
<tr>
<td>[44]</td>
<td>Apple (100 samples)</td>
<td>Starch pattern index</td>
<td>500–900</td>
<td>Natural light</td>
<td>PLSR</td>
<td>$R_v = 0.834$</td>
</tr>
<tr>
<td>[74]</td>
<td>Mango (78 samples)</td>
<td>Maturity</td>
<td>390.9–887.4</td>
<td>Clear sky conditions</td>
<td>CNN</td>
<td>$R_{CV} = 0.64$</td>
</tr>
</tbody>
</table>


Conducting research in natural orchard environments, which are incredibly complex, allows researchers to be closer to the actual growing environment than they would be in the laboratory. For instance, in an orchard environment, the intensity and angle of the sun’s rays change over time, and the wind can blow branches or leaves. In contrast, the halogen lamps used in studies performed under laboratory conditions are constant, and there is no wind. Field tests can consider more factors related to fruit growth, and real-time monitoring using portable spectroscopic instruments can capture changes in the spectral characteristics of fruit in different locations and at different times, more closely matching the environment in which the fruit is grown. The data obtained in this environment can be used to verify the accuracy of the laboratory model and adjusted and optimized according to the actual situation. Regarding current research, there are relatively few cases of fruit research in the orchard environment. Only for thin-skinned fruits, such as grapes, apples, mangoes, etc., and not for thick-skinned fruits such as citrus, have experiments and studies been carried out in depth [75].

As a result, there are still many uncharted research areas concerning fruits in orchard environments, including incredibly thick-skinned fruits such as citrus. Further research and experiments will help to reveal the complex mechanisms of fruit growth and development in orchard ecosystems, as well as provide a deeper understanding of optimizing orchard management and increasing fruit production by taking into account the complexity of the orchard environment.

4. Selection of Regions of Interest

In acquiring spectral data and images of a sample, the region to be processed is outlined from the image in boxes, circles, ellipses, irregular polygons, etc., known as the region of interest (ROI). ROI selection is mainly performed by various operators and functions commonly used in machine vision software, such as OpenCV, Matlab, etc., to derive the region of interest and to carry out the next step in processing the image. Figure 5a shows the region of interest obtained by threshold segmentation, which is the process of grayscaling the image to binarization [76]. Figure 5b shows the ROI selected by Tian P. et al. [77] using the manual frame selection method, where they averaged nine ROIs after framing them according to the three different parts of the mango.
Table 5 summarizes several typical methods of region of interest selection that have been used by researchers in spectroscopy-based nondestructive testing of fruit quality studies in the last five years, which will be discussed in the following two subsections.

4.1. Threshold Segmentation

In the acquisition process of hyperspectral images, the images captured contain not only sample information, but also information about the background. The background information is more complicated. For example, in the laboratory environment, it may be the desktop of the experimental platform, the ground, the belt of the conveyor belt, etc., and in the outdoor situation, the background is even more complicated, for example, the soil, the foliage, and the sky. Suppose all the information from the entire image is used as the spectral information of the sample. In that case, the background information would also be considered, causing significant interference in the results of the entire experiment. The threshold segmentation technique is one of the most classical and popular, and one of the most straightforward, image segmentation methods [84]. The key to this technique is finding the appropriate gray level threshold, which is usually selected based on the gray level histogram of the image. It involves dividing the gray level of an image into several parts using one or several thresholds, considering pixels belonging to the same part to be the exact object. This technique dramatically reduces the amount of data and simplifies...
the steps of analyzing and processing image information. A binary mask image is obtained by separating the samples from the background using thresholding. Then, this binary mask image is applied to obtain the ROI image.

Threshold segmentation is an important step in image analysis. If the segmentation is incorrect, background information may interfere with the ROI, or valuable sample data may be lost. There are more and more researchers using threshold segmentation to extract the regions of interest of samples in recent years, and for quantitative analysis, Shang M. et al. [45] chose an image with a wavelength of 1655.72 nm as the ROI image in a study of umbilical cord orange full-surface defect detection based on the hyperspectral on-line binning technique. The samples were segmented from the background using simple thresholding to obtain a binary mask image. Using the segmented image, an online sorting experiment was performed with 100% detection accuracy. Luo W. et al. [50] used hyperspectral imaging (380–1030 nm) to predict the SSC of Nanfeng mandarin, and the samples were separated from the background by threshold segmentation in order to reduce the background interference. After selecting the grayscale image at 850 nm to generate the mask image, all the original images were masked. Except for the Nanfeng mandarin calyx region, the other remaining parts were selected to be defined as ROI. The results showed that the bootstrapping soft shrinkage (BOSS)-competitive adaptive reweighted sampling (CARS)-partial least squares regression (PLSR) model predicted the best results, and could quickly and intuitively evaluate the intrinsic quality of Nanfeng honey orange.

In addition, in other fruits, experts have also conducted significant research. Çetin N. et al. [85] used hyperspectral technology and machine learning algorithms to evaluate the internal quality parameters of apple fruit. Firstly, the standard deviation of each pixel point was calculated, the samples were thresholded and segmented, and good segmentation results were obtained. The results showed that the potential use of hyperspectral imaging with artificial neural network and decision tree (DT) methods was more effective for hardness, while DT and multiple linear regression (MLR) were more effective for SSC. Li S. et al. [86], in a study on the ability of hyperspectral imaging combined with multiple regression modeling to detect loquat SSCs under small-sample conditions, utilized ENVI 5.1 software to detect loquat SSCs from the calibrated spectral images to select the entire loquat surface as the ROI and remove the background information. The results showed that a small sample size of hyperspectral images combined with an appropriate regression model can be used for nondestructive detection of loquat SSC. For qualitative analysis, Sharma S. et al. [87] obtained hyperspectral images of durian fruit flesh using a reflectance-based system for maturity grading assessment at full wavelength (900–1600 nm). The researchers extracted the ROI from each radio-corrected image with the threshold set to 0.5 and then applied a morphological filter to the binary image to extract the sample pixels. The results showed that support vector machine (SVM) and random forest (RF) classifiers can better utilize HSI to differentiate the ripening stages of durian flesh.

In summary, threshold segmentation provides excellent convenience for the acquisition of sample regions of interest, and it is computationally simple and fast, making it an ideal choice for large-scale samples. Additionally, the threshold segmentation method identifies key features in fruits, such as color and texture, which are crucial for studying ripeness, freshness, and intrinsic fruit quality [88]. However, two drawbacks must be considered. Firstly, most of the current literature focuses on segmenting the sample from the background, treating the entire sample as the region of interest. This approach leads to a significant increase in computational time and computational difficulty for subsequent reflectance calculations. Secondly, there are significant differences between the average reflectance of specific parts of particular samples and the whole, such as the reflectance of citrus fruits and the peduncle part. These differences may have a significant impact on the overall prediction or even lead to misclassification. Therefore, although the threshold segmentation method has obvious advantages in terms of acquiring the sample region of in-
terest and identifying key features, its computational time and difficulty need to be considered comprehensively when processing and analyzing large-scale samples [89]. At the same time, the significant differences in reflectance within specific samples should be handled with special care to avoid bias to the overall prediction, and these are areas that need to be focused on and improved in future research and applications.

4.2. Manual Selection

To solve the problem of the high computational time cost of threshold segmentation to extract the region of interest of the samples, some solutions have been proposed by experts in this field. Wang T. et al. [90], in a study on the determination of soluble solids content (SSC) in Korla pears by using the hyperspectral technique, used ENVI 5.3 to select 10 × 10 pixels of the hyperspectral image containing the spectral information of the entire surface of the balsam pear as the ROI, and the results showed that the SSC of Korla balsam pear was successfully predicted using a combination of clustering and edge influence analysis. Yu Y. et al. [91], to reduce the cost of nondestructive detection of the SSC of Korla balsam pear, developed a portable nondestructive SSC detector, which rotationally labeled three positions at every 120° on the equatorial line of the pear and obtained the local spectra as the ROI of the samples. The results showed that the developed SSC detector was also effective in the field. Gao Q. et al. [92], in a study on the evaluation of the soluble solids content, hardness index, and ripeness discrimination of Makino Begonia based on near-infrared hyperspectral imaging, a 20 pixel × 20 pixel area was selected on the upper surface of the fruits at intervals of 120°, with the stems as the center in the binomial image and the equator of the samples as a circle. A region was selected as the region of interest, three regions of interest were selected for each sample, and the average information of all regions of interest was regarded as the corresponding reflection spectrum of the sample. The results of this study showed that the NIR-HSI method is feasible for the quality evaluation of Makino begonia. Garillos-Manliguez C.A. et al. [93] estimated fruit ripeness by feature splicing of data acquired by visible and hyperspectral imaging systems. During ROI acquisition, nine equally spaced subregions (three in the middle portion, three near the vertices, and three near the shank) were obtained from the spectral image of the bounded ROI. Each subregion had a spatial resolution of 32 × 32 and a spectral resolution of 150. The results showed that, using multimodal inputs coupled with a powerful deep CNN model, it is possible to classify fruit ripeness even at the six-stage level of granularity. Mishra P. et al. [94], studied the improved prediction of the hardness of mango in the ripening stage using near-infrared spectroscopy supported by interval-partial least squares regression. The region of interest of mango samples was determined by selecting two different points which were distributed in the center of the two flanks, and the results showed that the predictions of the model provided the evolution of hardness throughout the ripening experiments.

In summary, the researchers utilized different shapes to select several regions of interest for samples in various locations. It was found that this method is feasible for selecting regions of interest to predict the component content and ripeness of the samples [95]. This method can effectively reduce the time cost. However, according to current studies, some experts randomly select several regions as the regions of interest of the sample. In contrast, others select several regions based on the parts of the sample. There is a certain degree of randomness and subjectivity in selecting regions of interest using the above two methods. Therefore, selecting representative and scientifically valid regions of interest has become an important focus and challenge in current research. With the continuous development of technology and methods, we can expect more breakthroughs in this field in the future. These advancements will provide more accurate and reliable methods and data support for studying citrus fruits and other fruits.
5. Selection of the Number of Samples and the Number of Spectral Features

Before modeling spectral analysis, the first task was to collect the sample data. The sample data included the sample’s raw spectral data and the sample’s component content information. The number of spectral features depends on the band range of the spectrometer used and the channel spacing, and in some cases, it is also related to the spectral noise [96]. The component content information of the samples depends on the number of samples collected in the experiment; however, the collection of the number of samples is affected by some external factors. Increasing the number of samples or spectral features does not necessarily result in better predictions. In naturally acquired samples, there are many duplicates and anomalous samples, which reduces the modeling speed and may decrease the accuracy of the model. Meanwhile, increasing the number of spectral features may lead to multicollinearity and the curse of dimensionality, which affect the model’s stability, computational efficiency, and generalization ability [97,98].

5.1. Quantitative Analysis

Researchers have conducted many studies on the quantitative analysis of fruits based on spectroscopy techniques using different numbers of samples and different number of spectral features, as shown in Table 6.

Table 6. Selection of the number of samples and the number of features in nondestructive testing by spectroscopy techniques.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Fruit (No. of Samples)</th>
<th>Sample Size</th>
<th>Quality Parameters</th>
<th>Wavelength Range (nm)</th>
<th>Number of Features</th>
<th>Data Analysis</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>[99]</td>
<td>Pear</td>
<td>65</td>
<td>Iron content</td>
<td>900–1700</td>
<td>228</td>
<td>PLSR</td>
<td>$R_p^2 = 0.753$</td>
</tr>
<tr>
<td>[100]</td>
<td>Dragon fruit, banana</td>
<td>100</td>
<td>SSC, TSS, TA, pulp content (PC)</td>
<td>400–1000</td>
<td>650</td>
<td>PLS</td>
<td>$R_p^2 = 0.59, 0.88$</td>
</tr>
<tr>
<td>[101]</td>
<td>Passion fruit</td>
<td>240</td>
<td>SSC</td>
<td>800–1098</td>
<td>298</td>
<td>PLSR</td>
<td>$R_p = 0.84, 0.91, 0.99$</td>
</tr>
<tr>
<td>[102]</td>
<td>Pomelo</td>
<td>100</td>
<td>TSS, acidity detection</td>
<td>400–1700</td>
<td>1300</td>
<td>PLSR</td>
<td>$R_p^2 = 0.72, 0.55$</td>
</tr>
<tr>
<td>[103]</td>
<td>Lemon</td>
<td>70</td>
<td>TSS, TA</td>
<td>950–1700</td>
<td>212</td>
<td>PLSR</td>
<td>$R_p^2 = 0.84, 0.72$</td>
</tr>
<tr>
<td>[104]</td>
<td>Apple</td>
<td>663</td>
<td>SSC</td>
<td>590–1200</td>
<td>203</td>
<td>PLS</td>
<td>$R_p = 0.9808$</td>
</tr>
<tr>
<td>[79]</td>
<td>Pear</td>
<td>185</td>
<td>SSC, firmness, MC</td>
<td>833–2500</td>
<td>2073</td>
<td>LSSVM</td>
<td>$R_p = 0.880, 0.826, 0.872$</td>
</tr>
<tr>
<td>[105]</td>
<td>Greengage fruit</td>
<td>366</td>
<td>sugar content</td>
<td>400–1000</td>
<td>256</td>
<td>PLSR</td>
<td>$R_p = 0.793$</td>
</tr>
<tr>
<td>[106]</td>
<td>Apple</td>
<td>118</td>
<td>SSC</td>
<td>450–1100</td>
<td>650</td>
<td>MNLR</td>
<td>$R^2 = 0.953$</td>
</tr>
<tr>
<td>[107]</td>
<td>Kiwifruit</td>
<td>210</td>
<td>glucose</td>
<td>400–1000</td>
<td>600</td>
<td>MLP</td>
<td>$R_p = 0.934$</td>
</tr>
<tr>
<td>[108]</td>
<td>Apple</td>
<td>174</td>
<td>sugar content</td>
<td>900–1750</td>
<td>850</td>
<td>PLS</td>
<td>$R_p = 0.916667$</td>
</tr>
<tr>
<td>[109]</td>
<td>Citrus</td>
<td>116</td>
<td>SSC</td>
<td>600–950</td>
<td>350</td>
<td>PLS</td>
<td>$R_p = 0.987$</td>
</tr>
<tr>
<td>[110]</td>
<td>Kiwifruit</td>
<td>800</td>
<td>SSC</td>
<td>950–1650</td>
<td>434</td>
<td>PLSR</td>
<td>$R = 4.386$</td>
</tr>
<tr>
<td>[111]</td>
<td>Winter jujube</td>
<td>400</td>
<td>SSC</td>
<td>400–1000</td>
<td>600</td>
<td>SVR</td>
<td>$R^2 = 0.837$</td>
</tr>
<tr>
<td>[112]</td>
<td>Kiwifruit</td>
<td>120</td>
<td>SSC</td>
<td>600–1100</td>
<td>111</td>
<td>PLS</td>
<td>$R^2 = 0.81$</td>
</tr>
<tr>
<td>[113]</td>
<td>Nectarine</td>
<td>480</td>
<td>SSC</td>
<td>420–1000</td>
<td>580</td>
<td>LSSVM</td>
<td>$R_p^2 = 0.8146$</td>
</tr>
<tr>
<td>[114]</td>
<td>Peach</td>
<td>150</td>
<td>SSC</td>
<td>350–1150</td>
<td>360</td>
<td>PLS</td>
<td>$R_p = 0.819$</td>
</tr>
</tbody>
</table>

Within the field, spectral resolution is a critical parameter in spectral analysis that describes a spectrometer’s ability to distinguish and discriminate between neighboring wavelengths. Many studies have shown that choosing a higher spectral resolution usually results in better predictions. The reason for this is that the reflection information in each band within the wavelength range covered by the spectrometer may be correlated with the component content of the sample. Therefore, researchers tend to narrow the spectral acquisition interval to capture more spectral features. Higher resolution captures more subtle spectral information in the sample and enhances the prediction model’s accuracy and stability. However, this creates two problems. The first problem is that the time cost of acquisition is significantly increased, and the second problem is that the wavelength screening of spectral features becomes complicated at a later stage. Some spectral data dimensionality reduction methods have to be used [115]. At the same time, we have found that some researchers choose larger spectral resolutions. However, fewer spectral bands may not provide enough information, leading to a decrease in the predictive model’s performance [116]. Specifically, a larger spectral resolution means that measurements are made over a larger wavelength interval, which reduces the number of spectral data points collected. In this case, the spectrometer may not be able to capture essential details and characteristic information in the sample, making it difficult for the model to resolve and predict the component content of the sample accurately. In addition, the reduced amount of information may make the model more sensitive to noise in the data, further reducing the generalization ability and stability. Therefore, while a larger spectral resolution may simplify data processing, it typically limits the predictive effectiveness and reliability of the model.

Compared to the selection of the number of spectral features, there is no fixed range for selecting the number of samples [117]. In a study, the choice of sample number can vary from 65 to 800. More samples can cover a broader range of fruit species and variability, resulting in more comprehensive and generalizable findings. However, large sample sizes also increase the complexity of data processing and analysis, requiring more advanced statistical methods and computational techniques [118]. In contrast, smaller sample sizes make data processing and analysis simpler and allow for quicker initial conclusions. However, the limited number of samples may lead to unstable results and make it difficult to assess the performance of spectroscopic techniques under different circumstances accurately. Therefore, to find a suitable balance between time cost, data processing complexity, and detection accuracy, the numbers of spectral features and samples need to be rationally selected. Such a balance helps to improve the comprehensiveness and reliability of the study while enabling a more effective assessment of the performance of spectroscopic techniques in different situations.

5.2. Qualitative Analysis

In addition to the quantitative analysis of fruits, there is also a choice of sample size and the amount of spectral characterization data for expert studies in the qualitative analysis of fruits, which mainly involves damage detection and disease detection. In the damage detection study, Du X.-L. et al. [119] collected visible/short-wave near-infrared (Vis/SWNIR) diffuse reflectance spectra from 300–1150 nm for analysis. Impact tests were conducted on 840 peach samples using two drop heights of 30 cm and 60 cm to study the variation in total soluble solids (TSS), which showed an $R^2$ of 0.89, an RMSEP of 0.40, and a residual predictive deviation (RPD) of 2.94 for TSS. Lee W.-H. et al. [120] investigated a new NIR technique, ultra-near infrared spectroscopic imaging at 950–1650 nm, for the detection of bruise damage in pear subcutis, and a total of 14 samples were evaluated in this study. The results of the study showed that the optimal threshold band ratio was 92% accurate for the detection of pear bruises. In a disease study, Sankaran S. et al. [121] collected spectral reflectance data from 100 healthy and 93 Huanglongbing-infected citrus trees for 989 spectral features in the wavelength range of 350–2500 nm using a visible-
near-infrared spectral radiometer. This was used to investigate the feasibility of visible-
near-infrared spectroscopy for the field detection of Huanglongbing (HLB) in citrus orchards. The results of the study showed that, comparing different classifier models, quadratic discriminant analysis (QDA) consistently resulted in higher average overall classification accuracy with lower false positives and false negatives. Ghooshkhaneh N.G. et al. [122] used VIS-NIR spectroscopy (400–1100 nm) to identify endophytic fungal diseases in order to collect the reflectance spectra of 540 oranges in three regions: stem end, equatorial end, and stigma end. The results showed that the highest classification accuracy was obtained using a back-propagation neural network classifier with optimal wavelengths in the stigma-end region. Xie C. et al. [123] used spectral reflectance for detecting citrus black spot (CBS) symptoms, setting the number of early, late, and healthy samples as 163, 163, and 58, respectively, in the training set, and 80, 81, and 28, and the spectra were collected in the range of 451–1010 nm. The results showed that the model correctly classified the diseased samples with accuracies of 99.4% (using the full wavelength) and 100.0% (using selected wavelengths) in the diseased versus healthy group, and for the early samples (early stage of the disease), the correctly classified accuracies were 92.5% (using the full wavelength) and 93.8% (using the selected wavelengths). Tian X. et al. [124] used image processing methods such as PCA, pseudo-color image transformation technology, and improved watershed segmentation algorithms (IWSA). A total of 132 normal fruits and 168 infested fruits with finger gram spores were prepared, and the feasibility of rot detection was analyzed based on 325–1098 nm hyperspectral transmission images. The results showed that, for the validation set containing 84 decayed and 66 sound fruits, the success rates were 93% and 96% for decayed fruits, 94% for sound fruits, and 94% for all fruits, respectively.

To summarize, in the qualitative analysis of fruit damage and disease based on spectroscopic techniques, unlike quantitative analysis, researchers are faced with the problem of choosing the number of different types of samples because they need to differentiate between several different types of samples, such as early hard spots, cracked spots, poison spots, and healthy samples. This complicates the selection of the number of samples. In this area, we have found that two situations exist. One is where the number of different types of samples is the same, usually to reduce classification bias and improve classification accuracy. The other situation is where the numbers of different types of samples are different, usually because specific categories of samples are more readily available or more critical. However, such an inconsistency may lead to over-representation of spectral features in specific categories, affecting the accuracy and reliability of qualitative analyses. Therefore, rational selection of sample size is essential to improve the comprehensiveness and reliability of the study. If the problem of inconsistent sample size cannot be solved by increasing the sample size, the bias can be corrected by adjusting the classification algorithm. For example, strategies such as resampling methods (oversampling and undersampling), weighted classifiers, data enhancement techniques, integrated learning, and adjusting decision thresholds are used [125]. These methods can balance categories, reduce classification bias, and improve the accuracy and reliability of qualitative analysis [126].

6. Conclusions

Spectroscopy is a core tool for the non-destructive quality testing of fruits. Although there have been some studies on the collection of spectral information from fruit samples, there are still many challenges regarding dataset collection methods. This paper reviews the selection of experimental environments, the selection of regions of interest, and the selection of the number of samples versus the number of spectral features during the spectral dataset collection process for citrus and other fruits. The enormous annual production of fruits limits the possibility of practical application of spectroscopy techniques in orchards. To perform non-destructive quality inspection of fruits more effectively and to
improve the efficiency of orchard management, further research on spectroscopy-based inspection of citrus and other fruits in the orchard environment is essential. Overcoming this technical challenge will contribute to the sustainable development of the fruit industry and the optimization of agricultural production. The main findings and conclusions are as follows. 

(1) Spectroscopic techniques excel in detecting defects, quantitative analysis, and ripeness classification of fruits such as citrus in a laboratory setting, where researchers can acquire accurate data under highly controlled conditions and calibrate and optimize models with standard samples. However, fewer studies have been conducted in realistic orchard environments with complex research conditions, especially for thick-skinned fruits such as citrus. Further research could help us to understand the mechanisms of fruit growth in orchard ecosystems and optimize orchard management to improve yields.

(2) The threshold segmentation method is simple, fast, and suitable for large-scale sample processing and critical feature identification, but suffers from problems of computational time and difficulty, as well as the challenge of internal reflectance differences within the sample, affecting prediction accuracy. Shape selection methods can reduce the time cost, but there is randomness and subjectivity, and selecting representative and scientific regions of interest is still challenging. Therefore, future research needs to address the computational issues of threshold segmentation methods and find more scientific and representative methods of selecting regions of interest to improve the accuracy and reliability of fruit studies.

(3) The accuracy of spectral analysis depends on the choice of spectral resolution and sample size. Higher resolution provides more accurate results, but increases time costs and complexity. A smaller sample size simplifies the analysis, but may reduce the stability of the results. A balanced choice of spectral features and sample size can help to improve the reliability and comprehensiveness of a study.

Despite some challenges faced by spectroscopy in the field of citrus and other fruit dataset collection, it still has a broad scope for development. It has shown irreplaceable importance in the field of agriculture. In the future, spectroscopy will play a vital role in the study of citrus and other fruits or crops.

**Author Contributions:** Supervision, review and editing: J.L.; conceptualization, investigation, writing—original draft preparation, review and editing: Q.W.; review and editing: Y.W.; methodology, conceptualization: J.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We thank Fajun Miao and Qiyang Shui for their help with this study.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**


Agriculture 2024, 14, 977

21 of 23


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.