agronomy

MDPI

Article

Strawberry Water Content Estimation and Ripeness Classification Using Hyperspectral Sensing

Rahul Raj [1,2,*], Akansel Cosgun [2] and Dana Kulić [2]

[1] CSRE, Indian Institute of Technology Bombay, Mumbai 400076, India
[2] Electrical and Computer Systems Engineering, Monash University, Clayton, Melbourne, VIC 3800, Australia; akansel.cosgun@monash.edu (A.C.); dana.kulic@monash.edu (D.K.)
* Correspondence: rahul.rahulraj@monash.edu

Abstract: We propose data-driven approaches to water content estimation and ripeness classification of the strawberry fruit. A narrowband hyperspectral spectroradiometer was used to collect reflectance signatures from 43 strawberry fruits at different ripeness levels. Then, the ground truth water content was obtained using the oven-dry method. To estimate the water content, 674 and 698 nm bands were selected to create a normalized difference strawberry water content index. The index was used as an input to a logarithmic model for estimating fruit water content. The model for water content estimation gave a correlation coefficient of 0.82 and Root Mean Squared Error (RMSE) of 0.0092 g/g. For ripeness classification, a Support Vector Machine (SVM) model using the full spectrum as input achieved over 98% accuracy. Our analysis further show that, in the absence of the full spectrum data, using our proposed water content index as input, which uses reflectance values from only two frequency bands, achieved 71% ripeness classification accuracy, which might be adequate for certain applications with limited sensing resources.

Keywords: fruit water content; fruit ripeness; hyperspectral data; machine learning for fruit ripeness

Citation: Raj, R.; Cosgun, A.; Kulić, D. Strawberry Water Content Estimation and Ripeness Classification Using Hyperspectral Sensing. *Agronomy* **2022**, *12*, 425. https://doi.org/10.3390/agronomy12020425

Academic Editors: Jaume Arnó and José A. Martínez-Casasnovas

Received: 31 December 2021
Accepted: 5 February 2022
Published: 8 February 2022

1. Introduction

The strawberry fruit is a high-value crop all across the world [1], having anti-inflammatory properties [2] and packed with vitamin C, fiber, and antioxidants. Being a non-climacteric fruit, strawberries need to ripen on the plant [3]. If plucked before or after maturity, its sugar, acidity, and phenolic contents deviate from optimal values, resulting in a reduction in its economic, aromatic, and health-related values [4]. Thus, harvesting the strawberry fruits at the right ripeness stage is an important step of strawberry farming [5]. Australia produces more than 76,640 tons of strawberries per year, having an economic value of approximately 450 million dollars. Approximately 85% of the strawberries are consumed within Australia, and 15% are exported.

Fruits can be detected from canopy images, based on color and texture differences between fruits and other parts of the plants [6–10]. However, vision-based estimation of fruit ripeness can be challenging, as the color and texture of fruits get saturated after some stage of the ripeness, making it difficult to detect when the fruit over-ripens. Detection of fruit ripeness based on non-destructive methods is important for its shelf life and industrial production [11].

As a fruit ripens, the percentage of water content and other biochemical contents inside the fruit changes [3,12–14]. Akhtar et al. (2015) [15] showed that the best time to harvest a strawberry fruit is when the water content in it is 90.6%, thus under this definition unripe strawberries contain more water, while overripe strawberries contain less water than this value. In a study by Frenkel et al. (2012) [3], it was found that a decrease in strawberry fruit water content is an early event in fruit ripening. Quantifying this decline, they showed that strawberry water content less than ~88 mL per gram of fresh tissue indicates the onset of

ripening. This reduction in water content with the increase in ripeness level is due to a change in water conversion of starch into sugar [16]. Due to this conversion of starch into sugars, ripe strawberry fruits with relatively lower water content taste sweeter than unripe strawberry fruit [12]. Similarly, Fecka et al. (2021) [17] analyzed various chemical properties of strawberry fruits and found that anthocyanins and polyphenols increase, but citric and other carboxylic acids decrease as the fruit ripens. Pineli et al. (2011) [18] conducted a similar experiment on different strawberry cultivars and found increasing anthocyanin in strawberries as the fruit ripens. Along with anthocyanin, chlorophyll and carotenoid are also present in strawberry fruits. Their concentration changes as the fruit ripens [19]. Interestingly, chlorophyll decreases, but anthocyanin increases as the fruit ripens [19,20]. However, the carotenoid trend is unclear as some researchers have found it slightly increasing [21], but many others have found it slightly decreasing [19,22]. The complex relation between strawberry water content and other fruit pigments, polyphenols, and acids makes the remote estimation of these contents a difficult problem.

The non-destructive study of strawberry fruits has been a point of interest to many researchers. RGB cameras are often used for visually detecting the location of fruits in images [9,10]; however, we focus our attention to previous work that used hyperspectral data to estimate water content and fruit ripeness, as the data provide rich information about the fruit. ElMasry et al. (2007) [23] used hyperspectral imaging (400–1000 nm) to determine quality attributes of the strawberry fruit, including fruit moisture content using a Machine Learning (ML) model. This study used a partial least square model to identify 480, 528, 608, 753, 817, 939, and 977 nm data helpful in estimating fruit moisture content, and obtained a correlation coefficient of 0.92. In another study, Liu et al. (2014) [24] used multispectral images to classify unripe, ripe, and overripe strawberries. The authors used a partial least square algorithm, and 90% accuracy was achieved. Shao et al. (2020) [25] conducted a lab-based study to determine strawberry ripeness using 400–1000 nm hyperspectral data. A machine learning classification model was used, and 100% accuracy was achieved. As this study was lab-based, a uniform light source was used with a constant sensor view angle, resulting in high classification accuracy. In another study, Gao et al. (2020) [26] collected on-farm 400–1000 nm hyperspectral data to estimate strawberry ripeness. The authors used a deep learning approach on the 400–1000 nm hyperspectral images. First, a sequential feature selection algorithm was applied and, on the first three principal components, a convolutional neural network was implemented. The authors achieved 98.6% accuracy on the test dataset. In a study conducted with RGB image-based strawberry ripeness estimation, Indrabayu et al. (2019) [27] used a support vector machine model with a radial basis function kernel for the classification of unripe, partially ripe, and ripe strawberries. The authors achieved 85% accuracy on the test dataset.

In some recent studies, Zhou et al. (2021) [28] collected drone-based high-resolution RGB images and used a YOLOv3 deep learning algorithm to classify strawberries in three and seven ripeness stages. The training data were prepared by manually Labeling the ripeness stages of strawberries. The authors achieved a precision between 0.72 and 0.93. Similarly, Anraeni et al. (2021) [29] also used strawberries RGB images and implemented feature extraction and k-Nearest Neighbor to identify the ripeness of strawberries. The authors achieved 85% classification accuracy. In another study, Devassy and George (2021) [30] used successive projections algorithms on 400–1000 nm hyperspectral data. Authors identified 804, 552, and 667 nm as most useful for firmness detection of the strawberry fruits. Firmness is one of the aspects of fruit ripeness.

2. Method Overview

The majority of non-destructive estimation of strawberry fruit water content research focused on data-dependent machine learning models. Moreover, complex relations among various pigments and corresponding fruit water content have not been explored for making a conceptual fruit water content estimation model. In this study, a conceptual, non-destructive strawberry fruit water content estimation model is developed for fruit water

content estimation and machine learning models are implemented for ripeness measurement. In the 350–2500 nm range narrow-band and pure pixel spectral signatures were collected from strawberry fruits at different ripeness levels. Post data collection, the strawberry fruits were cut into small pieces and oven-dried to obtain the ground-truth fruit water content.

We propose two contributions: water content estimation (Section 4) and ripeness classification (Section 5). For water content estimation, based on light reflectance characteristics of water and other strawberry fruit pigments, two specific frequencies in the spectral data were identified for creating a normalized index. The index was later used in a conceptual model based on the Beer–Lambert law [31] for the development of our water content estimation model. Further, the complete spectroradiometer data were empirically analyzed, and results are critically discussed. For ripeness classification, we train three different machine learning algorithms (Decision Tree, Support Vector Machine, and Multilayer Perceptron) with an ablation study on input data and report on classification accuracy. The overview of our approach is shown in Figure 1.

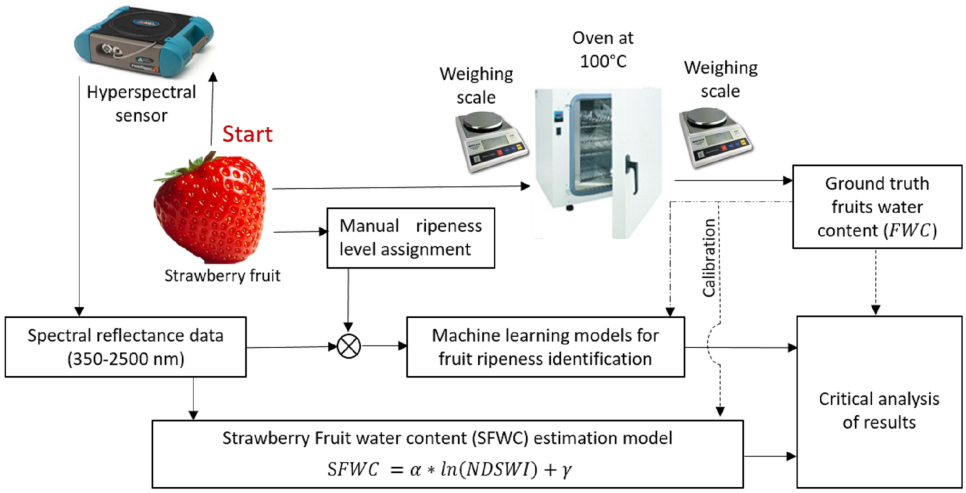

Figure 1. The framework of the strawberry fruit water content estimation and ripeness classification. The strawberry fruit water content estimation and ripeness classification framework. In the SFWC block, α is the gain, γ is the offset, and NSDSWI stands for Normalized Difference Strawberry Water Content Index.

3. Data Collection

In this section, we describe our procedure on how the strawberries were sourced, the hyperspectral data were collected, the water content ground truth was measured, and the preprocessing we apply to the hyperspectral data for removing noise.

3.1. Strawberry Acquisition

First, strawberries were acquired from local retail stores and from pot-grown strawberry plants. Some rotten fruits were discarded and only normal-looking strawberries were selected for the study. A total of 43 strawberries were used for data collection, out of which 2 were from the strawberry plants and 41 were purchased from a retail store. All the purchased strawberries were kept at atmospheric temperature before recording spectral signatures from them. Out of the 43, 5 strawberries had soft skin and dark red color, 18 had glossy red color skin, and 20 had skin color with different combinations of white, pink, and red. The color- and skin tone-based identification were done based on the work of Mohamed et al. (2019) [32]. Upon further visual inspection, we manually identified 28 samples as ripe, 13 as unripe, and 2 as over-ripe. The data collection was carried out on

25 March 2020, at Monash University Campus, outdoors under natural sunlight and clear sky conditions. The strawberries used in this study are shown in Figure 2.

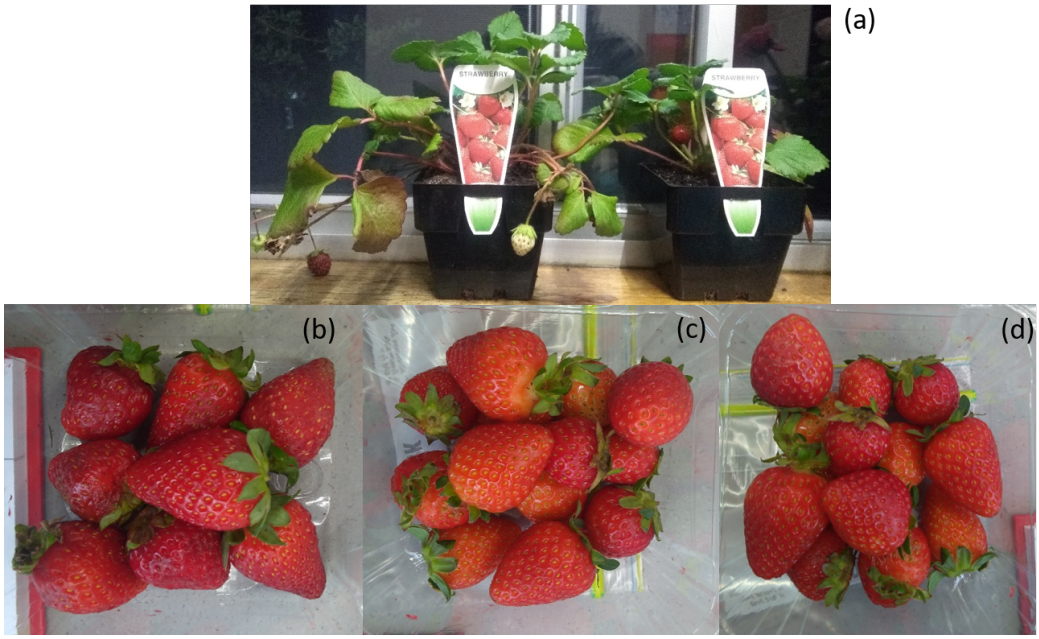

Figure 2. Strawberries used in this research: (**a**) pot grown strawberries; (**b**–**d**) retail store acquired strawberries.

3.2. Hyperspectral Data Collection

Reflectance spectral signatures of the individual strawberry fruits were collected using a 350–2500 nm range-calibrated ASD Fieldspec spectroradiometer (Analytical Spectral Devices Inc., Boulder, CO, USA), as shown in Figure 3A. The bandwidth of each band was 1 nm, thus the total number of bands was 2150. A 1° field of view lens cap was used for data collection to ensure recording of pure-pixel reflectance from strawberry fruits. The sensor was kept around 5–8 cm from the strawberry fruits, and the scanning direction of the sensor was maintained orthogonal to the direction of the incident and reflected sunlight. This was done to eliminate specular radiation and to only record diffused sunlight reflected from the sample, as that contains information about the internal structure and other chemical properties of the samples. From each of the fruits, two spectral signatures were collected, resulting in a total of 86 recorded spectra. After collecting data from each strawberry, a barium sulfate reference panel was used to collect the reference spectra to minimize the effect of any change in solar illumination. After each data capture, the strawberries were packed into an airtight plastic zip bag and a unique code was assigned to each of the samples and sent to the laboratory for further analysis.

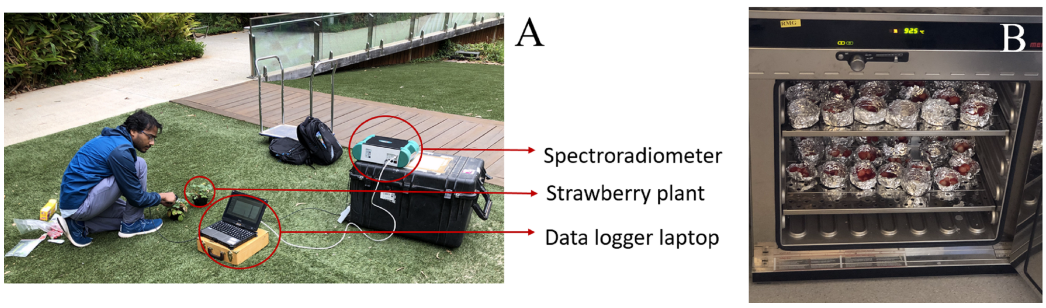

Figure 3. (**A**) Data collection from ASD spectroradiometer under natural sunlight conditions. (**B**) After the spectroradiometer data collection, the samples were weighed and placed in an oven with marked aluminum containers.

3.3. Water Content Measurement

In the laboratory, the strawberries were taken out from the zip-bags one by one, cut into pieces, weighed, and put into a pre-weighed, pre-coded aluminum container. The weight of the strawberry fruits at this stage was considered as fresh weight. As shown in Figure 3B, all 43 samples were then put in an industrial oven at 100 °C for 48 h to dry the samples thoroughly. Two sample weights were also recorded every 6–10 h, and it was found that the dry weights saturated after 30 h. All the samples were kept in the oven for 18 more hours to make sure they dried completely. The sample-filled containers were weighed again, and the weight of the aluminum container was subtracted from the total weight. The obtained weight was considered as the dry weight. The difference between fresh and dry weight strawberries was used to calculate the water content of the strawberry fruits as

$$\%water\_content = 100 \times (w_{fresh} - w_{dry})/(w_{fresh}) \quad (1)$$

where w_{fresh} is the fresh weight and w_{dry} is the dry weight. The measured water content of all strawberry samples was within the 87% to 94% range.

3.4. Data Preprocessing

The fruit reflectance spectra were collected using the spectroradiometer, and respective fruit water content obtained from the oven-dried method was documented in a CSV file. The 350–2500 nm spectroradiometer raw data contain high-frequency noise as shown in Figure 4. Thus, a Savitzky–Golay filter was run on each spectrum to remove the high-frequency noise. Figure 5a shows one raw reflectance spectra collected from the spectroradiometer and its smoothed version. Figure 5b shows the part of electromagnetic spectra highlighting the difference between the raw and smooth spectra. The bands around 1400 nm (1351–1420 nm), 1900 nm (1801–2040 nm), and 2400 nm (2300–2500 nm) were removed from analysis as the data in those wavelength ranges were noise because of the atmospheric opaqueness of the atmosphere. These noisy bands have been removed from the final data set, and only 1640 bands were used for further analysis.

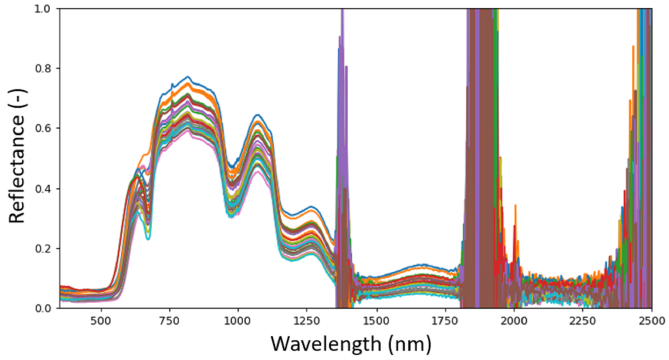

Figure 4. Raw hyperspectral reflectance spectra of some strawberry fruit samples.

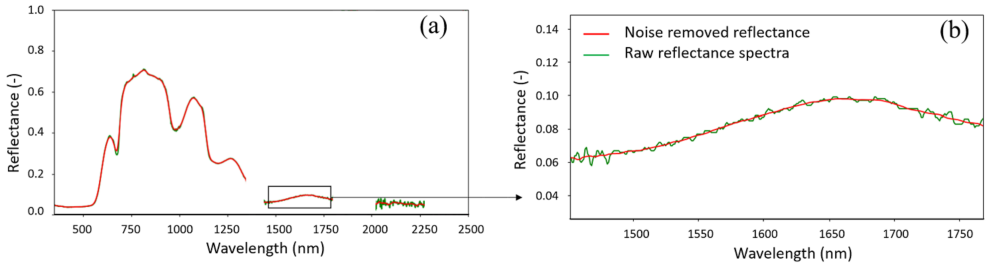

Figure 5. (**a**) Raw hyperspectral reflectance spectra (green) of a strawberry fruit and its smoothed version (red). (**b**) Zooming in the reflectance spectra indicated within a box in panel (**a**), which highlights the effect of the Savitzky–Golay filter at relatively less noisy parts of the spectra.

4. Strawberry Water Content Estimation

4.1. Strawberry Fruit Water Content Metric

We propose an index which is a function of reflectance values at two specific wavelengths:

$$Index_{mn} = (R_m - R_n)/(R_m + R_n) \quad (2)$$

where R_m and R_n denote the reflectance at band m and n, respectively. It is common practice in the literature to develop such indices for estimating various crop properties and identifying the best bands by exhaustively searching all combinations [33–35]. We adopt the same idea and search for all possible combinations of pairs of wavelengths from 400 nm to 2300 nm frequencies. In total, more than 1.2 million two-band normalized difference index combinations were created using Equation (2). The coefficient of determination of each of these indices was calculated with the ground-truth strawberry water content, and a correlation heatmap was created, which is shown in Figure 6. The coefficient of determination was found positive for all the indices.

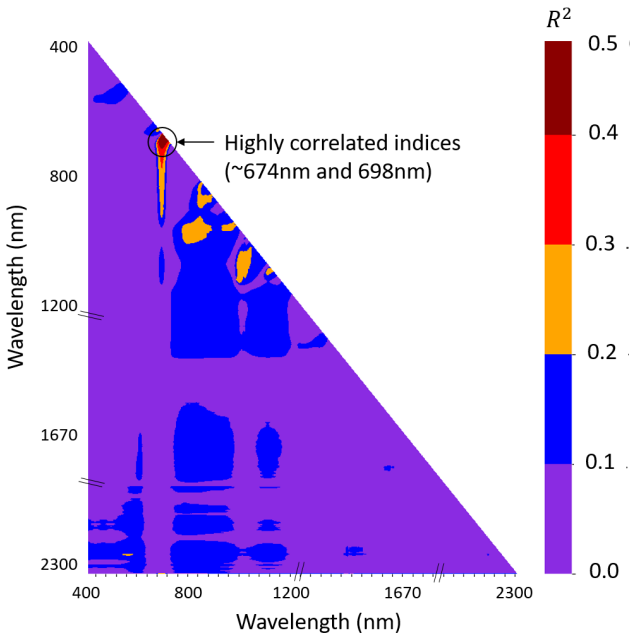

Figure 6. Correlation heatmap of two-band normalized difference index with the strawberry water content. Red- and orange-shaded areas highlight wavelengths involved in most informative indices. The line break symbols on the wavelength axes indicate the omission of noisy wavelengths, as explained in Section 3.4.

It can be seen in Figure 6 that the highest correlation values occur in a small region in the spectrum. To determine the exact bands for the index, we tap into the existing literature. The light absorptance characteristics of water and various pigments have been shown to have the same absorption coefficient for different chlorophyll, carotenoids, and anthocyanin contents in the wavelength interval of 670–750 nm [36–39]. We thus consider the 670–750 nm electromagnetic spectrum region as a suitable frequency interval for analyzing the recorded spectroradiometer data, as the effects due to change in strawberry water content will not be affected due to change in other pigments in the fruit. Our data also confirm that the indices with highest correlations lie within this frequency band. Within this frequency band, we picked two specific narrow-band wavelength frequencies to develop our model:

- 674 nm: A water absorption wavelength has previously been identified at 698 nm by Braun and Smirnov (1993) [40] and validated by Raj et al. (2021) [41] on spectroradiometer and drone-based hyperspectral data. Thus, to create a reliable strawberry

water content index, the 698 nm wavelength should be connected with any other wavelength in the 670–750 nm range.

- 698 nm: We look for the band that has the maximum correlation with our chosen band of 674 nm. This frequency is far from the water absorption wavelength, sufficiently inside the boundary of the 670–750 nm region, and falls within the visible electromagnetic spectrum range, making it more accessible for many researchers.

We thus define Normalized Difference Strawberry Water Content Index (NDSWI) as

$$NDSWI = (R_{698} - R_{674})/(R_{698} + R_{674}) \quad (3)$$

where R_x is the reflectance at x nm wavelength. The normalized difference index helps in scaling the information relative to reflectance values at both wavelengths. The value of NDSWI is directly proportional to the water content in the strawberry fruit. The concentration of water can be related to the absorbance of electromagnetic radiation as per Beer–Lambert law [31,42]. We use a modified Beer–Lambert law to create the Strawberry Fruit Water Content (SFWC) metric:

$$SFWC = \alpha \times ln(NDSWI) + \gamma \quad (4)$$

where α is the gain, and γ is the offset. As we show in Equation (4), the SFWC is a function of the natural logarithm of NDSWI. The values of the constants α and γ are initially unknown and need to be empirically determined from a sample dataset. We calculate these constants by randomly selecting 60% of the collected data using the k-fold cross-validation approach. A total of 100 times random sampling was done in the k-fold cross-validation approach leading to 100 different values of α and γ. The k-fold cross-validation approach is used to reach optimal values of the measuring variable. As per Equation (4), a curve was fitted through the sample data points using the generalized reduced gradient nonlinear method [43]. The obtained values of α and γ were 0.038 ± 0.005 and 0.98 ± 0.04, respectively. Here, 0.005 and 0.04 are errors associated with the mean values of α and γ values, respectively. This may lead to the change in slope and offset by 13% and 4%, respectively, contributing to $\pm 3.5\%$ fruit water content in the worst case. The obtained numeric equation for the strawberry fruit water content (gram/gram) model is thus

$$SFWC = 0.038ln(NDSWI) + 0.98 \quad (5)$$

4.2. Water Content Estimation

The developed model, as shown in Equation (5), was evaluated on the test data. The test data had 33 samples which the model has not seen before. The NDSWI values were fed as input to the developed model, and strawberry water content was obtained for the 33 samples. Figure 7a shows the trend of the developed model on the test data. The estimated water content was correlated with ground truth strawberry water content and a correlation coefficient of 0.82 and Root Mean Square Error (RMSE) of 0.0092 g/g were obtained. Figure 7b shows the scatter plot between ground-truth and model-estimated strawberry fruit water content. The RMSE of the build model is less than one percent of the change. This shows that the model is able to distinguish strawberry fruits with different water content with accuracy of 1% change in value. The direct correlation between NDSWI and ground truth water content gave a correlation coefficient of 0.66 and RMSE of 0.012 g/g (shown in Figure 8a), which was also the best among all possible normalized indices correlation with strawberry water content as shown in Figure 6. This indicates that the developed SFWC conceptual model is an improvement over the linear regression model between NDSWI and the water content of strawberry fruit.

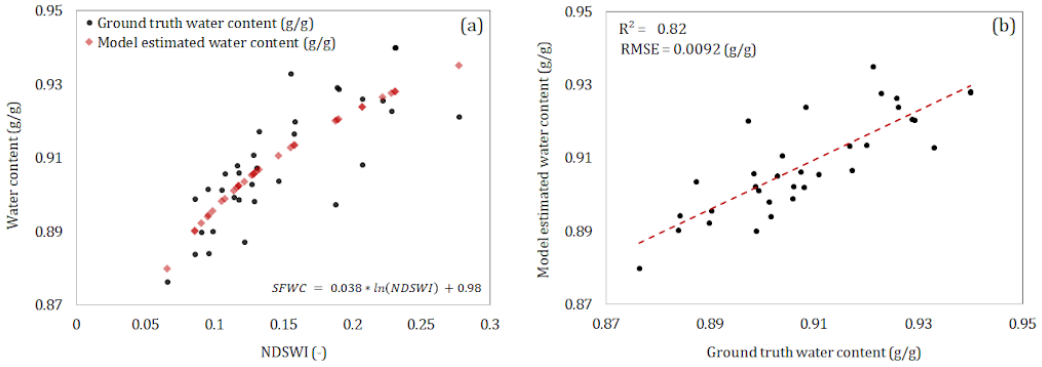

Figure 7. (**a**) Scatter plot of strawberry fruit water content model (SFWC) result and ground truth strawberry fruit water content, against normalized difference strawberry water content index (NDSWI). (**b**) Scatter plot between ground truth and estimated strawberry water content.

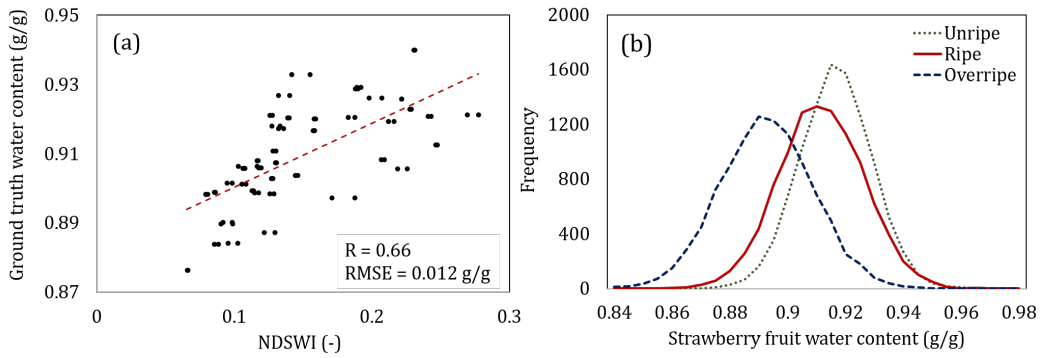

Figure 8. (**a**) Scatter plot between NDSWI and ground truth strawberry water content. (**b**) Histogram of unripe, ripe, and overripe strawberry fruits water content values.

As identified by Akhtar et al. (2015) [15], in this research we also identified that the strawberries having glossy red color skin had water content of approximately 90–91%. In the literature discussed in the introduction section, it was found that strawberries with glossy red color skin mostly represent ripe strawberries. The strawberries with soft skin and of dark red color (indicating overripe condition) most had a water content of around 88–89%. In contrast, strawberries of skin color with different combinations of white, pink, and red (indicating partially ripe or unripe condition) had a water content of approximately 92%. The three categories of strawberries water content are statistically plotted in Figure 8b. Figure 8b is a histogram plot drawn by generating 10,000 normal random values based on the mean and standard deviation of the samples classified in the three categories (unripe, ripe, and overripe) during data collection. It can be seen that the water content of overripe and partially ripe strawberries differs by 3% for the majority of the data. The water content of ripe strawberries falls between the ripe and unripe water content values. Given that our model is capable of identifying 1% water content change in the strawberries, it can be used for ripeness estimation of the strawberries. However, the water content threshold values for identifying fruit ripeness need to be identified for the different verities of fruits. Note that the 43 strawberries used in the study may belong to mixed varieties, which could have influenced the threshold values. Additionally, as some of the strawberries were purchased from a retail store they may thus lose some water content over time, leading to deviation in identified water content threshold values. Researchers working in this field may consider this a future research topic and consider conducting fruit species-based analysis.

5. Strawberry Ripeness Classification

In this section, we train various machine learning models for classifying each strawberry as one of the three classes (ripe, unripe, and overripe), with different inputs. The data

and code used for this paper are made publicly available (https://github.com/acosgun/strawberry_ripeness, accessed on 7 February 2022). The classification accuracy results are presented in Table 1.

Table 1. Strawberry ripeness classification accuracy for different models and different input data.

Input	Input Size	ML Method	Ripe %	Unripe %	Overripe %	Overall %
NDSWI	1	Decision Tree	76.6	55	100	71.2
NDSWI	1	SVM (rbf)	94.1	22.4	0	68
NDSWI	1	MLP	98.2	9.2	0	66.6
NDSWI+WC	2	Decision Tree	92	75.6	100	87.4
NDSWI+WC	2	SVM (rbf)	100	0	0	65
NDSWI+WC	2	MLP	100	0	0	65
670–750 nm band	81	Decision Tree	77.9	53.3	69	70.1
670–750 nm band	81	SVM (rbf)	93.1	26	0	68.4
670–750 nm band	81	MLP	90.7	29.2	0	67.8
Spectrum	1640	Decision Tree	97.7	90.4	88	93.8
Spectrum	1640	SVM (linear)	99.7	94.9	100	98.2
Spectrum	1640	MLP	97.9	92.7	99	96.4
Spectrum+WC	1641	Decision Tree	95.6	90.7	87	93.7
Spectrum+WC	1641	SVM (linear)	99.8	94.9	100	98.3
Spectrum+WC	1641	MLP	97.2	93.3	98	96.1

Three machine learning models were implemented using scikit-learn Python software library:

1. Decision Tree: Gini impurity is used for measuring split quality metric and the best split is chosen at each node.
2. Support Vector Machine (SVM): Different kernels (linear, polynomial, rbf, sigmoid) were separately tried and the the model that yielded the highest test accuracy was chosen. The linear kernel was used for models that takes the spectrum as input, and the rbf kernel was used for models that takes NDSWI as input.
3. Multi Layer Perceptron (MLP): A fully connected neural network with 4 hidden layers of sizes 80, 40, 20, 10 using adam optimizer, a batch size of 50, and ReLu activation function was trained for 1000 epochs without early stopping.

Combinations of three kinds of inputs are used:

1. Normalized Difference Strawberry Water Content Index (NDSWI): The metric we propose in Equation (3), that uses reflectance at 698 nm and 674 nm wavelengths.
2. Measured Water Content (WC): The ground-truth water content percentage obtained from the oven-dry method as explained in Section 3.3.
3. Spectrum: Hyperspectral data with preprocessing, as explained in Section 3.4.

Because of the very small number of data samples (86 data points in total), we employed a cross-validation technique for splitting train and test sets to ensure the resulting numbers are robust. Specifically, we employed a stratified 5-fold cross-validation method (which results in an 80/20% train and test split), repeated 25 times with random seeds. That means that each model (that corresponds to a row in Table 1) was trained 125 times with randomly chosen train/test sets, and the mean classification accuracy is reported in Table 1. Stratification seeks to ensure that each fold is representative of the class proportions of the data. This is especially useful for a very small dataset like ours.

As it can be seen in Table 1, the best performing ML model was SVM trained on all available data (spectrum and water content). When we do not assume access to ground truth water content, the overall classification accuracy was 98.2%, with perfect accuracy

on overripe, 99.7% accuracy on ripe, and 94.9% accuracy on unripe strawberries. These results show that our ripeness classifier performed better than Indrabayu et al. [27] and Shao et al. [25], on par with Gao et al. [26] who used hyperspectral imaging, and slightly worse than Liu et al. [24] who reported perfect accuracy. While Liu et al. [24]'s results are impressive, we note two important differences between our experiment setup and theirs: (1) Our data were collected outdoors while Liu et al. [24] conducted the measurements in a controlled environment, and (2) they performed leave-one-out cross-validation once, while we perform a stratified k-fold cross-validation method, with the mean classification accuracy computed over 25 repeated, randomly initiated trials.

Our analysis further shows that, in the absence of the full spectrum data, using NDSWI as input which uses reflectance values from only two specific frequencies, the decision tree model achieved a respectable 71.2% overall classification accuracy. This model actually performed slightly better than the model that uses 81 reflectance values in the 670–750 nm band (which was identified as an informative band as discussed in Section 4.1). This observation validates the utility of the proposed NDSWI index for strawberry ripeness estimation, in addition to water content estimation. If the ground truth water content measurement is also available in addition to the NDSWI, the decision tree model can achieve a relatively high accuracy (87.4%). This is a design implication for researchers who might want to estimate ripeness with limited sensing or computing resources.

When we compare the performance of the ML models, decision trees performed the best when the input consisted of only a few numbers, whereas SVM models performed best with larger input sizes, i.e., when larger portions of the spectrum data was part of the input. SVM is a commonly used method in the literature, which often performs well on hyperspectral data [24–27]; however, to our knowledge, decision trees have not been utilized for ripeness estimation. Given the superior results of decision tress with small input sizes, we encourage other researchers in this field to try out this learning method as well. MLP models did not perform well (in line with Liu et al. (2014) [24]), which might be explained by the input data type and the dataset size. Not only are neural networks known to excel with structured data such as images, but they are also known as data-hungry models, which likely rendered MLP's models unsuitable for our problem. Another observation is the volatility of the accuracy numbers between different models for classifying overripe strawberries—while some models achieved perfect accuracy, some had zero accuracy. This is an artifact of the small dataset size—as there were only four measurements that were taken from overripe strawberries in our dataset. Moreover, the manual ripeness level classification of fruit may add some errors in the process. We believe that a larger dataset would not only eliminate such artifacts but would likely increase the accuracy numbers across the board.

6. Conclusions

In this research, a new approach has been developed to estimate strawberry fruit water content for the identification of the ripeness level using a non-destructive method. One narrowband normalized difference index was created using spectroradiometer data based on strawberry fruit water and pigment content light absorption characteristics. The two narrow hyperspectral bands (around 1 nm bandwidth) were centered around 674 and 698 nm. The normalized difference index was used as an input to a modified Beer–Lambert law equation to create a conceptual model for fruit water content estimation. The model gave a correlation coefficient of 0.82 and RMSE of 0.0092 g/g. The obtained fruit water content can be used to identify the ripeness levels of the fruit. Accordingly, this model can be used in farms to determine the harvesting time of the fruits in real-time, which will help in minimizing the loss of fruit quality. Our experiments into ripeness classification showed that using the full spectrum rather than a couple of frequencies performed better by achieving over 98% accuracy. Finally, our proposed water content index proved to be useful for both water content estimation and ripeness classification, and can be useful when sensing is limited to a few bands instead of the full spectrum.

Author Contributions: Conceptualization, R.R., A.C. and D.K.; methodology, R.R. and A.C.; software, R.R.; validation, R.R.; formal analysis, R.R.; resources, R.R. and A.C.; data curation, R.R.; writing—original draft preparation, R.R.; writing—review and editing, R.R., A.C. and D.K.; visualization, R.R.; supervision, A.C. and D.K.; project administration, A.C.; funding acquisition, A.C. and D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Bosch Agriculture Technology LaunchPad Program.

Data Availability Statement: The data used in this research is available at https://github.com/acosgun/strawberry_ripeness, accessed on 7 January 2022.

Acknowledgments: The authors would like to acknowledge Alicia Quinn for providing the necessary resources to conduct lab experiments for this research. Special thanks to Sunny Goondram for helping with the hyperspectral data collection and oven drying process.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Simpson, D. The economic importance of strawberry crops. In *The Genomes of Rosaceous Berries and Their Wild Relatives*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 1–7.
2. Giampieri, F.; Forbes-Hernandez, T.Y.; Gasparrini, M.; Alvarez-Suarez, J.M.; Afrin, S.; Bompadre, S.; Quiles, J.L.; Mezzetti, B.; Battino, M. Strawberry as a health promoter: An evidence based review. *Food Funct.* **2015**, *6*, 1386–1398. [CrossRef] [PubMed]
3. Frenkel, C.; Hartman, T.G. Decrease in fruit moisture content heralds and might launch the onset of ripening processes. *J. Food Sci.* **2012**, *77*, S365–S376. [CrossRef] [PubMed]
4. Jediyi, H.; Naamani, K.; Elkoch, A.A.; Dihazi, A.; El Fels, A.E.A.; Arkize, W. First study on technological maturity and phenols composition during the ripeness of five Vitis vinifera L grape varieties in Morocco. *Sci. Hortic.* **2019**, *246*, 390–397. [CrossRef]
5. Sharma, S.; Singh, K. Harvesting. In *Strawberries*; CRC Press: Boca Raton, FL, USA, 2019; pp. 399–402.
6. Puttemans, S.; Vanbrabant, Y.; Tits, L.; Goedemé, T. Automated visual fruit detection for harvest estimation and robotic harvesting. In Proceedings of the 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), Oulu, Finland, 12–15 December 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6.
7. Lamb, N.; Chuah, M.C. A strawberry detection system using convolutional neural networks. In Proceedings of the 2018 IEEE International Conference on Big Data (Big Data), Seattle, WA, USA, 10–13 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 2515–2520.
8. Yu, Y.; Zhang, K.; Yang, L.; Zhang, D. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN. *Comput. Electron. Agric.* **2019**, *163*, 104846. [CrossRef]
9. Goondram, S.; Cosgun, A.; Kulic, D. Strawberry Detection using Mixed Training on Simulated and Real Data. In Proceedings of the International Conference on Digital Image Computing: Techniques and Applications (DICTA), Melbourne, Australia, 29 November–2 December 2020.
10. Sa, I.; Ge, Z.; Dayoub, F.; Upcroft, B.; Perez, T.; McCool, C. Deepfruits: A fruit detection system using deep neural networks. *Sensors* **2016**, *16*, 1222. [CrossRef]
11. Wei, X.; Liu, F.; Qiu, Z.; Shao, Y.; He, Y. Ripeness classification of astringent persimmon using hyperspectral imaging technique. *Food Bioprocess Technol.* **2014**, *7*, 1371–1380. [CrossRef]
12. Sturm, K.; Koron, D.; Stampar, F. The composition of fruit of different strawberry varieties depending on maturity stage. *Food Chem.* **2003**, *83*, 417–422. [CrossRef]
13. Díaz-Pérez, J.C.; Muy-Rangel, M.D.; Mascorro, A.G. Fruit size and stage of ripeness affect postharvest water loss in bell pepper fruit (*Capsicum annuum* L.). *J. Sci. Food Agric.* **2007**, *87*, 68–73. [CrossRef]
14. Misron, N.; Aliteh, N.A.; Harun, N.H.; Tashiro, K.; Sato, T.; Wakiwaka, H. Relative estimation of water content for flat-type inductive-based oil palm fruit maturity sensor. *Sensors* **2017**, *17*, 52. [CrossRef]
15. Akhtar, I.; Rab, A. Effect of fruit ripening stages on strawberry (Fragarta X Ananassa. Duch) fruit quality for fresh consumption. *J. Agric. Res.* **2015**, *53*, 213–224.
16. Forney, C.; Kalt, W.; McDonald, J.E.; Jordan, M. Changes in strawberry fruit quality during ripening on and off the plant. In Proceedings of the International Postharvest Science Conference Postharvest 96 464, Taupo, New Zealand, 4–9 August 1996; pp. 506–506.
17. Fecka, I.; Nowicka, A.; Kucharska, A.Z.; Sokół-Łętowska, A. The effect of strawberry ripeness on the content of polyphenols, cinnamates, L-ascorbic and carboxylic acids. *J. Food Compos. Anal.* **2021**, *95*, 103669. [CrossRef]
18. Pineli, L.d.L.d.O.; Moretti, C.L.; dos Santos, M.S.; Campos, A.B.; Brasileiro, A.V.; Córdova, A.C.; Chiarello, M.D. Antioxidants and other chemical and physical characteristics of two strawberry cultivars at different ripeness stages. *J. Food Compos. Anal.* **2011**, *24*, 11–16. [CrossRef]
19. Woodward, J. Physical and chemical changes in developing strawberry fruits. *J. Sci. Food Agric.* **1972**, *23*, 465–473. [CrossRef]

20. Yoshida, Y.; Koyama, N.; Tamura, H. Color and anthocyanin composition of strawberry fruit: Changes during fruit development and differences among cultivars, with special reference to the occurrence of pelargonidin 3-malonylglucoside. *J. Jpn. Soc. Hortic. Sci.* **2002**, *71*, 355–361. [CrossRef]
21. Delgado-Pelayo, R.; Gallardo-Guerrero, L.; Hornero-Méndez, D. Carotenoid composition of strawberry tree (*Arbutus unedo* L.) fruits. *Food Chem.* **2016**, *199*, 165–175. [CrossRef]
22. Zhu, H.; Chen, M.; Wen, Q.; Li, Y. Isolation and characterization of the carotenoid biosynthetic genes LCYB, LCYE and CHXB from strawberry and their relation to carotenoid accumulation. *Sci. Hortic.* **2015**, *182*, 134–144. [CrossRef]
23. ElMasry, G.; Wang, N.; ElSayed, A.; Ngadi, M. Hyperspectral imaging for nondestructive determination of some quality attributes for strawberry. *J. Food Eng.* **2007**, *81*, 98–107. [CrossRef]
24. Liu, C.; Liu, W.; Lu, X.; Ma, F.; Chen, W.; Yang, J.; Zheng, L. Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit. *PLoS ONE* **2014**, *9*, e87818. [CrossRef]
25. Shao, Y.; Wang, Y.; Xuan, G.; Gao, Z.; Hu, Z.; Gao, C.; Wang, K. Assessment of strawberry ripeness using hyperspectral imaging. *Anal. Lett.* **2020**, *54*, 1547–1560. [CrossRef]
26. Gao, Z.; Shao, Y.; Xuan, G.; Wang, Y.; Liu, Y.; Han, X. Real-time hyperspectral imaging for the in-field estimation of strawberry ripeness with deep learning. *Artif. Intell. Agric.* **2020**, *4*, 31–38. [CrossRef]
27. Indrabayu, I.; Arifin, N.; Areni, I.S. Strawberry Ripeness Classification System Based On Skin Tone Color using Multi-Class Support Vector Machine. In Proceedings of the 2019 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 24–25 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 191–195.
28. Zhou, X.; Lee, W.S.; Ampatzidis, Y.; Chen, Y.; Peres, N.; Fraisse, C. Strawberry Maturity Classification from UAV and Near-Ground Imaging Using Deep Learning. *Smart Agric. Technol.* **2021**, *1*, 100001. [CrossRef]
29. Anraeni, S.; Indra, D.; Adirahmadi, D.; Pomalingo, S.; Sugiarti; Mansyur, S.H. Strawberry Ripeness Identification Using Feature Extraction of RGB and K-Nearest Neighbor. In Proceedings of the 2021 3rd East Indonesia Conference on Computer and Information Technology (EIConCIT), Surabaya, Indonesia, 9–11 April 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 395–398.
30. Devassy, B.M.; George, S. Estimation of strawberry firmness using hyperspectral imaging: A comparison of regression models. *J. Spectr. Imaging* **2021**, *10*. [CrossRef]
31. Swinehart, D.F. The beer-lambert law. *J. Chem. Educ.* **1962**, *39*, 333. [CrossRef]
32. Mohamed, I.; Williams, D.; Stevens, R.; Dudley, R. Strawberry ripeness calibrated 2D colour lookup table for field-deployable computer vision. In *Proceedings of the IOP Conference Series: Earth and Environmental Science*; IOP Publishing: Bristol, UK, 2019; Volume 275, p. 012003.
33. Thenkabail, P.S.; Smith, R.B.; De Pauw, E. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* **2000**, *71*, 158–182. [CrossRef]
34. Fava, F.; Colombo, R.; Bocchi, S.; Meroni, M.; Sitzia, M.; Fois, N.; Zucca, C. Identification of hyperspectral vegetation indices for Mediterranean pasture characterization. *Int. J. Appl. Earth Obs. Geoinf.* **2009**, *11*, 233–243. [CrossRef]
35. Delegido, J.; Verrelst, J.; Meza, C.; Rivera, J.; Alonso, L.; Moreno, J. A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *Eur. J. Agron.* **2013**, *46*, 42–52. [CrossRef]
36. Feild, T.S.; Lee, D.W.; Holbrook, N.M. Why leaves turn red in autumn. The role of anthocyanins in senescing leaves of red-osier dogwood. *Plant Physiol.* **2001**, *127*, 566–574. [CrossRef] [PubMed]
37. Gitelson, A.A.; Merzlyak, M.N.; Chivkunova, O.B. Optical properties and nondestructive estimation of anthocyanin content in plant leaves. *Photochem. Photobiol.* **2001**, *74*, 38–45. [CrossRef]
38. Hughes, N.M.; Smith, W.K. Attenuation of incident light in Galax urceolata (Diapensiaceae): concerted influence of adaxial and abaxial anthocyanic layers on photoprotection. *Am. J. Bot.* **2007**, *94*, 784–790. [CrossRef]
39. Merzlyak, M.N.; Chivkunova, O.B.; Solovchenko, A.E.; Naqvi, K.R. Light absorption by anthocyanins in juvenile, stressed, and senescing leaves. *J. Exp. Bot.* **2008**, *59*, 3903–3911. [CrossRef]
40. Braun, C.L.; Smirnov, S.N. Why is water blue? *J. Chem. Educ.* **1993**, *70*, 612. [CrossRef]
41. Raj, R.; Walker, J.P.; Vinod, V.; Pingale, R.; Naik, B.; Jagarlapudi, A. Leaf water content estimation using top-of-canopy airborne hyperspectral data. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102393. [CrossRef]
42. Nikzad-Langerodi, R.; Zellinger, W.; Saminger-Platz, S.; Moser, B.A. Domain adaptation for regression under Beer–Lambert's law. *Knowl.-Based Syst.* **2020**, *210*, 106447. [CrossRef]
43. Lasdon, L.S.; Waren, A.D.; Jain, A.; Ratner, M. Design and testing of a generalized reduced gradient code for nonlinear programming. *ACM Trans. Math. Softw. (TOMS)* **1978**, *4*, 34–50. [CrossRef]