

Hyperspectral Imaging and Convolutional Learning for Sugariness Prediction: Example of *Syzygium Samarangense*

Chih-Jung Chen

National Yang Ming Chiao Tung University

Yung-Jhe Yan

National Yang Ming Chiao Tung University

Chi-Cho Huang

Fengshan Tropical Horticultural Experiment Branch

Jen-Tzung Chien

National Yang Ming Chiao Tung University

Chang-Ting Chu

National Yang Ming Chiao Tung University

Je-Wei Jang

National Yang Ming Chiao Tung University

Tzung-Cheng Chen

Feng Chia University

Shiou-Gwo Lin

National Taiwan Ocean University

Ruei-Siang Shih

National Yang Ming Chiao Tung University

Mang Ou-Yang (✉ oym@nctu.edu.tw)

National Yang Ming Chiao Tung University

Research Article

Keywords: Hyperspectral image, Convolutional neural network, Feedforward neural network, Sugariness, *Syzygium samarangense*

Posted Date: July 2nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-663300/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Hyperspectral Imaging and Convolutional Learning for Sugariness Prediction: Example of *Syzygium samarangense*

Chih-Jung Chen¹, Yung-Jhe Yan¹, Chi-Cho Huang², Jen-Tzung Chien¹, Chang-Ting Chu¹, Je-Wei Jang¹, Tzung-Cheng Chen³, Shiou-Gwo Lin⁴, Ruei-Siang Shih¹, Mang Ou-Yang^{1*}

1. National Yang Ming Chiao Tung University, Taiwan
2. Fengshan Tropical Horticultural Experiment Branch, Taiwan
3. Feng Chia University, Taiwan
4. National Taiwan Ocean University, Taiwan

Abstract

Sugariness is one of the most important indicators to measure the quality of *Syzygium samarangense*, which is also known as the wax apple. In general, farmers used to measure sugariness by testing the extracted juice of the wax apple products. Such a destructive way to measure sugariness is not only labor-consuming but also a wasting of products. Therefore, non-destructive and quick techniques for measuring sugariness would be significant for wax apple supply chains. Traditionally, the non-destructive method to predict the sugariness or the other indicators of the fruits was based on the reflectance spectra or Hyperspectral Images (HSIs) using linear regression such as Multi-Linear Regression (MLR), Principal Component Regression (PCR), and Partial Least Square Regression (PLSR), etc. However, these regression methods are usually too simple to precisely estimate the complicated mapping between the reflectance spectra or HSIs and the sugariness. This study presents the deep learning methods for sugariness prediction using the reflectance spectra or HSIs from the bottom of the wax apple. A non-destructive imaging system fabricated with two spectrum sensors and light sources is implemented to acquire the visible and infrared lights with a range of wavelengths. In particular, a specialized Convolutional Neural Network (CNN) with hyperspectral imaging is proposed by investigating the effect of different wavelength bands for sugariness prediction. In the experiments, the ground-truth value of sugariness is obtained from a commercial refractometer. The experimental results show that using the whole band range between 400 nm and 1700 nm achieves the best performance in terms of Brix error. CNN models attain the Brix error of ± 0.552 , smaller than ± 0.597 using Feedforward Neural Network (FNN). Significantly, the CNN's test results show that the minor error in Brix interval 0-10°Brix and 10-11°Brix are ± 0.551 and ± 0.408 , these results indicate that the model would have the capability to predict if sugariness is below 10°Brix or not, which would be similar to the human tongue. These results are much better than ± 1.441 and ± 1.379 by using PCR and PLSR, respectively. Moreover, this study provides the test error in each Brix interval within one °Brix, and the results show that the test error is varied considerably within different Brix intervals, especially on PCR and PLSR. On the other hand, FNN and CNN obtain robust results in terms of test error.

Keywords: Hyperspectral image · Convolutional neural network · Feedforward neural network · Sugariness · *Syzygium samarangense*

Introduction

The fruit's quality grading is one of the significant factors in processing agricultural products. A proper procedure for grading is beneficial to preservation and transportation and increasing the value of the products. The precise classification and reasonable price make customers willing to purchase the agriculture products. However, the farmers and dealers subjectively grade the quality and prices of fruits by the product's size, appearance, color or sweetness, etc., not precisely because the criterion may not be unified. *Syzygium samarangense*, also known as wax apple, for example, the farmers usually determine the sweetness of the wax apple by the color of the peel. With more scarlet the color is, the more sugar content may be contained within the fruit. However, such judgment is not only time-consuming but also easily miscalculates the quality of the products. For more precise detection, auxiliary instruments, such as the °Brix meter, have already been developed for detecting the sugar content in the

pulp. This kind of °Brix meter often utilizes light refraction through the liquid to observe and detect a small detecting error.

Nevertheless, the drawback of the refracted meter is the procedure that needs to destroy the product itself to extract the juice for detection. This invasive method is not only time-consuming but also impractical for real applications in the agricultural field. Consequently, the non-destructive sweetness detection technique will become a crucial issue for measuring the quality of wax apple fruit.

In general, fruit's non-destructive detecting methods are implemented by analyzing the reflectance spectra or the Hyperspectral Image (HSI) of the surface of the fruit. Rather than ordinary RGB images, HSI would have more features and be an ideal material to analyze. For instance, previous works presented that hyperspectral imaging is more suitable than conventional RGB imaging for evaluating mushroom quality¹, and the advantages of analysis on HSI are shown^{23,4}. In the early researches, the analyzing methods including Multiple Linear Regression (MLR), Principal Component Regression (PCR), and Partial Least Square Regression (PLSR). Such as the soluble solids content of peaches could be determined using MLR³. The experiments apply PLSR on the reflectance of apples to detect the firmness⁶, the quality of gala apples could be measured by detecting the soluble solids and firmness using PCR⁷. The concentration of anthocyanins, polyphenols, and sugar of grapes was determined using PLSR⁸ and a previous work uses Independent Component Analysis (ICA) and PLSR to quantify sugar content in wax apple using spectra data with NIR bands (600-1098nm)¹⁵. Moreover, some proposed researches focus on the nutrients in the wax apple, such as proving that the total anthocyanin content (TAC) and total phenolic compounds (TPC) were able to be detected by near-infrared spectroscopy¹¹.

In recent years, deep learning is prevalent in developing non-destructive technique of determination of quality from various fruits, such as the ripeness of strawberry could be analyzed with CNN and SVM models on HIS data⁹, and the maturity of citrus could be estimated with fluorescence spectroscopy data by performing the regression with a CNN model¹⁰. Furthermore, deep learning is also broadly applied in the actual farm field, such as one of the researches proposed research uses R-CNN based model to detect and count passion fruit in orchards¹², another implements CNN models on wheat yield forecasting¹³, and the other achieves high accuracy on grape bunch segmentation using deep neural networks¹⁴.

Those experimental results indicate the potential possibility of using the spectral convolutional neural network to detect the fruit quality. Therefore, take the fundamental requirements from the agricultural field and the spectral technique's opportunity into consideration.

Data Preparation and Evaluation Process

Figure 1 shows the workflow of the entire evaluation, which consists of five steps. First, the HSIs of the bottom part of the wax apple were acquired properly and preprocessed as a 1-dimensional array for *t*-distributed Stochastic Neighbor Embedding (*t*-SNE) data visualization, FNN, PCR, and PLSR models, the 3-dimensional matrix for the 2D-CNN models. Second, *t*-SNE was used to visualize the dataset to see the distribution of data. Third, the preprocessed data was split into training, validation, and test datasets. Fourth, the FNN and CNN models were trained with training and validation datasets.

Furthermore, to review if the clustering results are better than the input samples after the deep learning process, *t*-SNE was applied on the layer's outputs before the regression output layer. Finally, each model was evaluated with the test dataset, and the error was assessed for different models. Further information on HSIs preprocessing and modeling in this research will be introduced in the following few parts of this section.

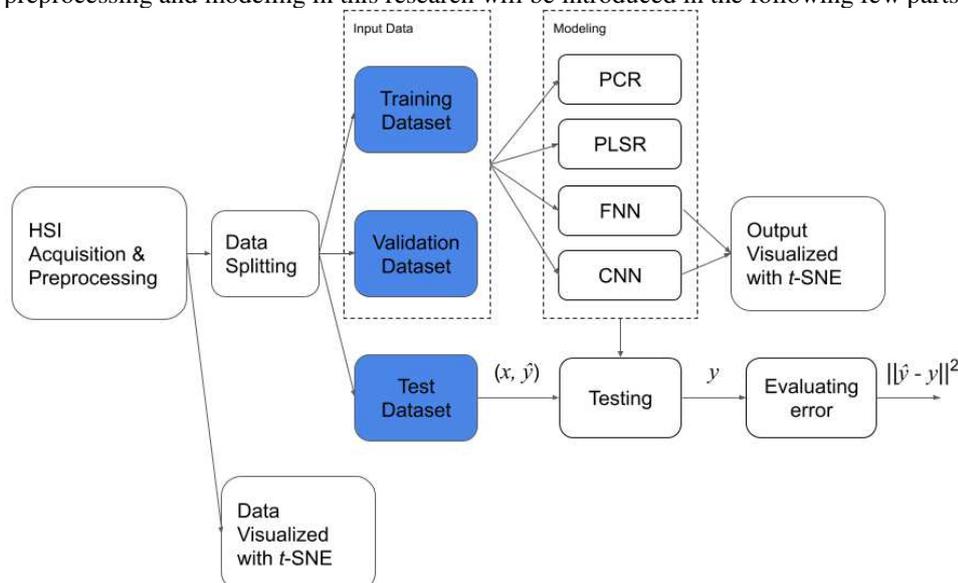


Figure 1. Overall workflow for evaluating the related methods in sugariness prediction.

Samples Preparation

The most important step is preparing the samples, so 136 wax apple samples (Tainung No.3, Sugar Barbie) were collected from the orchards located in Liouguei, Jiadong, Meishan, and Fengshan, Taiwan. The farmers of these wax apples are directed by experts from Fengshan Tropical Horticultural Experiment Branch (FTHEB). The wax apple samples were transported directly from different farms to a lab located at FTHEB. After wax apple samples were transferred to the lab, they were gathered and recovered to room temperature because the samples were refrigerated shipping. This step can make sure that there is no water drop on the surface of the wax apple, which may cause errors for reflectance measurement. When the samples recovered to room temperature, they were cut into small pieces. By the experience of farmers and the professional experts from FTHEB, there is more sugar content in the down part of the wax apple, and the study referred¹⁵ focuses on analyzing the reflectance data of the down part of the wax apple. Therefore, the pieces from the down part of the wax apple were picked up and stabbed on the blackboard to collect the HSIs data.

HSIs Data and Sugariness Measurement

The hyperspectral data were collected in a dark room to reduce the measurement noise with the equipment made our team¹⁶, which was applied to the experiment. This hyperspectral system integrated two different sensors for detecting the visible (VIS) light and short-wave infrared (SWIR). By the designed mechanism, the sample spectra can be collected with a spectral range from 400 nm to 1700 nm simultaneously. This device can obtain the complete spectral information from the surface of the wax apple, and the total size of the HSI is ($W \times L \times C$), where W is width, L is the length, and C is the number of hyperspectral bands. Compared with spectroscopy, this device can additionally analyze the changes of spectra.

After the scanning process was done, the Brix value was measured by crushing the samples to extract the juice to measure the Brix value. The average sugar content was 11.49 °Brix, and the standard deviation was 1.945° Brix. The sugar content calculated with a commercial refractometer ATAGO PAL-1, these recorded values will be the labeled targets for training deep neural networks.

The preprocessing of HSIs was crucial for model training. Therefore, the HSIs were calibrated with spatial and White/Dark light calibration and smooth the waveform with a Savitzky-Golay filter¹⁷. For data augmentation, a procedure was designed to randomly sample the HSIs of pieces of wax apple into the smaller cube with size $20 \times 20 \times C$, and the number of smaller cubes each piece can take was according to its original size (width around 30 to 90 pixels and length around 90-150 pixels), 1034 pieces in total. The 3-D HSIs datasets will be used to train CNN models, and for t -SNE data visualizing and training on FNN models, PCR and PLSR, the 3-D datasets were averaged to 1-D datasets by their width and length. The entire data augmentation workflow is shown in the Figure 2 Note that in this study, the analytical strategy was focused on spectral analysis. Therefore, the proposed data augmentation method nullified the influence of spatial features of the data.

In this study, data sets were re-sampled for data balancing. The data sets were classified into five groups according to its ground-truth label, including "under Brix 10", "Brix 10-11", "Brix 11-12", "Brix 12-13" and "Brix above 13" and the number of spectra data in each group was sampled as identical as possible.

After data balancing, the small cube (size $20 \times 20 \times 1367$) HSIs datasets will be quadrupled into four data sets with different band range, including 400-1700nm (each HSI with size $20 \times 20 \times 1367$), 400-1000nm (each HSI with size $20 \times 20 \times 1053$), 400-700nm (each HSI with size $20 \times 20 \times 575$) and 900-1700nm (each HSI with size $20 \times 20 \times 424$) because this study aims to evaluate which band range has better correlation toward making sugariness prediction. These 3-D cubes datasets will be used to train CNN models, and for t -SNE data visualizing and training on FNN models, PCR and PLSR, the 3-D datasets are averaged to 1-D datasets by their width and length, so there had also four data sets including 400-1700nm (each array with size 1×1367), 400-1000nm (each array with size 1×1053), 400-700nm (each array with size 1×575) and 900-1700nm (each array with size 1×424). After the data sets are prepared, they were randomly sampled into "training", "validation", and "test" for modeling. All the number of data and size of each set is as shown in Table 1, and the Figure 3 shows the entire workflow for data sets preparation.

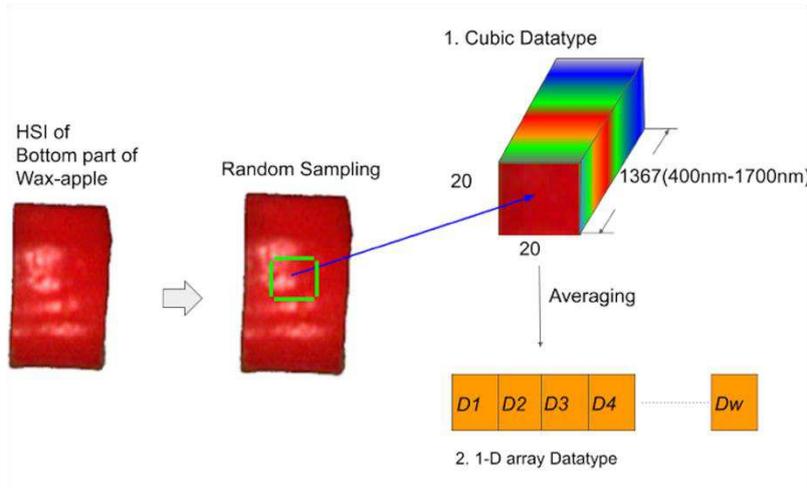


Figure 2. The workflow shows that the data augmentation for generating 3-D hyperspectral images, and 1-D spectrum averages. The smaller HSI will be randomly sampled, and the number of the smaller HIS depends on the size of the wax apple sample. Because this study aims to analyze the spectral features of HSIs, the spatial features would not be considered in the data augmentation process.

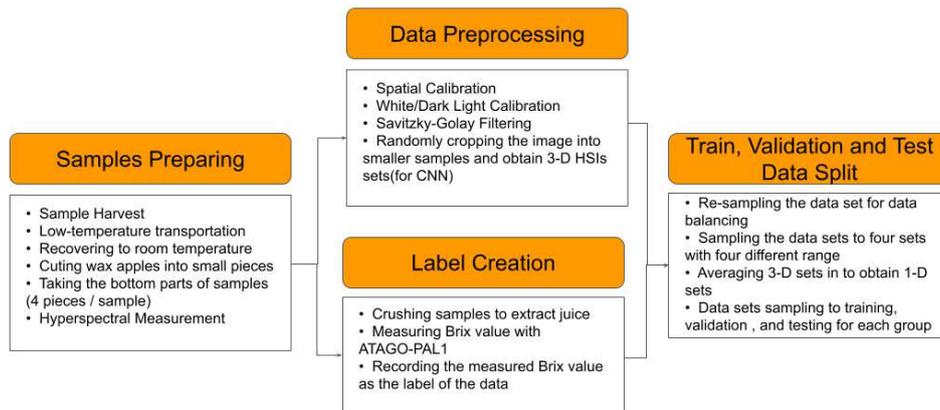


Figure 3. The workflow is designed for preparing experimental datasets. First, the samples are transferred to the lab under low-temperature transportation. Second, after the samples are recovered to room temperature, the samples are cut into small pieces for HSIs scanning and juice extracting. Third, the HSIs are obtained by the imaging system in a dark room. Fourth, the small samples are crushed by a small juicer to extract the juice for measuring the Brix value, which is the data label, with a refractometer ATAGO-PAL1. Fifth, the HSIs are calibrated with spatial calibration and white/dark light calibration, then Savitzky-Golay Filtering is applied to smooth the waveform. Seventh, each of the HSI is randomly cropped into smaller samples for data augmentation. Seventh, the data sets are divided into five groups, including "under Brix 10", "Brix 10-11", "Brix 11-12", "Brix 12-13", and "Brix above 13" depending on the Brix value. The numbers of data in groups are similar for data balancing. Eighth, the 3-D HSIs data sets are sampled into four datasets with different band range, including 400-1700nm (each HSI with size $20 \times 20 \times 1367$), 400-1000nm (each HSI with size $20 \times 20 \times 1053$), 400-700nm (each HSI with size $20 \times 20 \times 575$) and 900-1700nm (each HSI with size $20 \times 20 \times 424$). Then, the 3-D datasets are averaged along their width and length to obtain 1-D datasets. Therefore, there had also four data sets, including 400-1700nm (each array with size 1×1367), 400-1000nm (each array with size 1×1053), 400-700nm (each array with size 1×575), and 900-1700nm (each array with size 1×424). Finally, all the data sets are sampled into "training", "validation", and "test" for modeling.

Table 1. The table shows all the information of data sets, each 3-D and 1-D data sets consist of four different wavelength range spectra datasets where each spectra dataset is composed of five groups of random samples for Brix under 10, Brix 10-11, Brix 11-12, Brix 12-13 and Brix above 13. The M denotes the mean Brix value, and Std denotes the standard deviation of samples in each group.

Range	Size	Groups	Number of samples		
			Training	Validation	Test
400-1700nm					

3-D HSIs dataset	20×20 $\times 1367$	Brix under 10 (M:8.654, Std:0.823)	131	44	43	
		Brix 10-11 (M:10.448, Std:0.251)	97	33	32	
		Brix 11-12 (M:11.487, Std:0.231)	131	44	43	
		Brix 12-13 (M:12.526, Std:0.272)	131	44	43	
		Brix above 13 (M:14.016, Std:0.888)	131	44	43	
400-1000nm	20×20 $\times 1053$	Brix under 10 (M:8.654, Std:0.823)	131	44	43	
		Brix 10-11 (M:10.448, Std:0.251)	97	33	32	
		Brix 11-12 (M:11.487, Std:0.231)	131	44	43	
		Brix 12-13 (M:12.526, Std:0.272)	131	44	43	
		Brix above 13 (M:14.016, Std:0.888)	131	44	43	
400-700nm	20×20 $\times 575$	Brix under 10 (M:8.654, Std:0.823)	131	44	43	
		Brix 10-11 (M:10.448, Std:0.251)	97	33	32	
		Brix 11-12 (M:11.487, Std:0.231)	131	44	43	
		Brix 12-13 (M:12.526, Std:0.272)	131	44	43	
		Brix above 13 (M:14.016, Std:0.888)	131	44	43	
900-1700nm	20×20 $\times 424$	Brix under 10 (M:8.654, Std:0.823)	131	44	43	
		Brix 10-11 (M:10.448, Std:0.251)	97	33	32	
		Brix 11-12 (M:11.487, Std:0.231)	131	44	43	
		Brix 12-13 (M:12.526, Std:0.272)	131	44	43	
		Brix above 13 (M:14.016, Std:0.888)	131	44	43	
1-D spectra data	400-1700nm	1×1367	Brix under 10 (M:8.654, Std:0.823)	131	44	43
			Brix 10-11 (M:10.448, Std:0.251)	97	33	32
			Brix 11-12 (M:11.487, Std:0.231)	131	44	43
			Brix 12-13 (M:12.526, Std:0.272)	131	44	43
			Brix above 13 (M:14.016, Std:0.888)	131	44	43
400-1000nm	1×1053	Brix under 10 (M:8.654, Std:0.823)	131	44	43	
		Brix 10-11 (M:10.448, Std:0.251)	97	33	32	
		Brix 11-12 (M:11.487, Std:0.231)	131	44	43	
		Brix 12-13 (M:12.526, Std:0.272)	131	44	43	
		Brix above 13 (M:14.016, Std:0.888)	131	44	43	
400-700nm	1×575	Brix under 10 (M:8.654, Std:0.823)	131	44	43	
		Brix 10-11 (M:10.448, Std:0.251)	97	33	32	
		Brix 11-12 (M:11.487, Std:0.231)	131	44	43	
		Brix 12-13 (M:12.526, Std:0.272)	131	44	43	
		Brix above 13 (M:14.016, Std:0.888)	131	44	43	
900-1700nm	1×424	Brix under 10 (M:8.654, Std:0.823)	131	44	43	
		Brix 10-11 (M:10.448, Std:0.251)	97	33	32	
		Brix 11-12 (M:11.487, Std:0.231)	131	44	43	
		Brix 12-13 (M:12.526, Std:0.272)	131	44	43	
		Brix above 13 (M:14.016, Std:0.888)	131	44	43	

Data Visualization with t -SNE

The t -Distributed Stochastic Neighbor Embedding (t -SNE) is a nonlinear machine learning algorithm proposed by van der Maaten and Hinton¹⁸. It is a popular manifold learning algorithm that aims to preserve the feature information when high dimensional data are visualized as low dimensional data. In t -SNE, the goal is to make the conditional probability (equation 1) of high (p_{ij}) and low (q_{ij}) dimensional Euclidean distances as close as possible by minimizing the cost function (equation 2), which is the sum of Kullback-Leibler divergences over all data points.

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|x_k - x_i\|^2 / 2\sigma_i^2)}, \quad q_{ij} = \frac{(1 + \|y_i - y_j\|^2 / 2\sigma_j^2)^{-1}}{\sum_{k \neq i} \exp(-\|x_k - x_i\|^2 / 2\sigma_i^2)} \quad (1)$$

$$C = \sum_i KL(P_i \| Q_i) = \sum_i p_{j|i} \log \frac{p_{j|i}}{q_{j|i}} \quad (2)$$

And because σ_i is varied for every data point, how σ_i is selected is here in accordance with a fixed perplexity

$$Perp(P_i) = 2^{H(P_i)} \quad (3)$$

where $H(P_i)$ is the Shannon Entropy P_i measured in bits

$$H(P_i) = - \sum_j (p_{j|i} \log_2 p_{j|i}) \quad (4)$$

In this study, data visualization could help to do preliminary analysis. Therefore t -SNE was used to scale the 1-D data sets (obtained from the preprocessing stage) from size $N \times D$ to $N \times 2$, where N was the number of samples 1034 and D was the number of data points of whole the bands, then the N data points were visualized on the 2-D plane. The entire workflow is shown in the Figure 4.

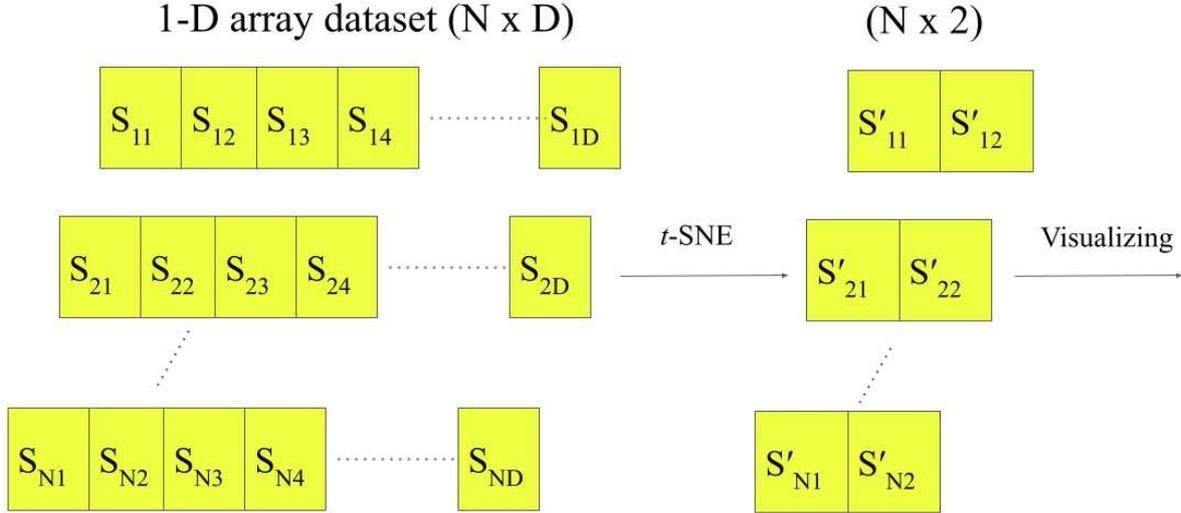


Figure 4. The workflow shows how this study implements t -SNE. The 1-D arrays with spectra data(S) corresponding to wavelength are averaged from 3-D HSIs datasets along width and length and reduced to size, then visualized on the 2-D plane.

Furthermore, in order to visualize the distribution of the data after the deep learning process to make a contrast, t -SNE was also used to evaluate the clustering of the outputs before the last regression output layer of proposed deep learning models. Figure 5. shows the workflow of t -SNE to present learning results. First of all, X was the input dataset with total N wax apple samples, and each data in X was fed into the proposed deep learning models to obtain output from the layer which was fully-connected to the last regression output layer. Second, the N 1-D arrays with each size $1 \times Q$, where Q denoted as the number of the layer before the last output layer, were composed of each the output obtained by feeding each data in X. Finally, the matrix with size $N \times Q$ was scaled to matrix $N \times 2$, which was able to visualize in the 2-D plane.

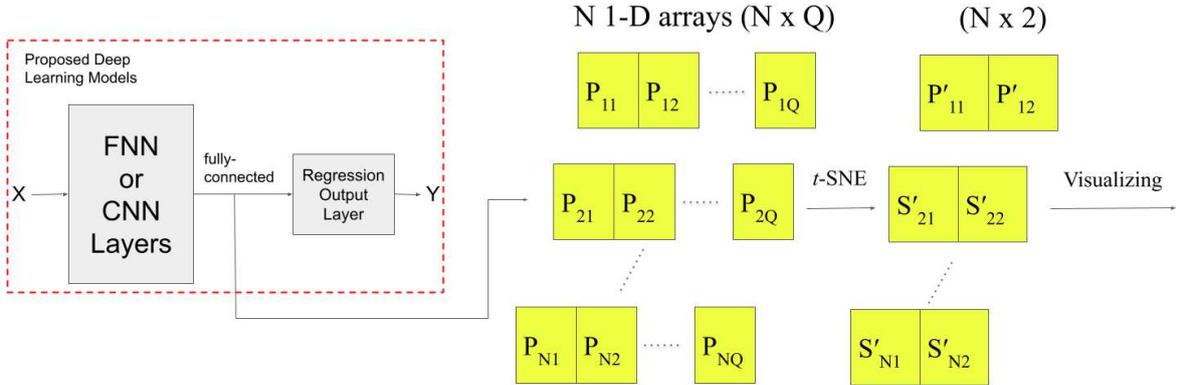


Figure 5. The workflow shows how t -SNE evaluates the learning results. First, the matrix with the size is obtained by feeding the dataset X to proposed deep learning models and collecting the layer's outputs before the last regression output layer. Then the matrix is scaled to N data points to visualize the t -SNE result.

Hyperspectral and Convolutional Modeling

Two linear regression techniques and two deep learning modeling techniques were implemented on our datasets. The linear algorithms Principal Component Regression (PCR) and Partial Least Square Regression (PLSR) were applied to compare with non-linear deep learning techniques on 1-D spectrum array datasets in this research. Respectively, the proposed deep learning models were Feedforward Neural Network (FNN) and Convolutional Neural Network (CNN). These two techniques will be used on 1-D spectra datasets and 3-D HSIs datasets.

Principal Component Regression and Partial Least Square Regression

PCR and PLSR were applied to our dataset to compare with deep learning methods. PCR is a method based on the Principal Component Analysis (PCA) and Multi-Linear Regression (MLR) approach. The main idea of the PCA is to find out the projection vector v^* that can maximize the variation of the projected data and can be denoted as equation 5, where the M is the covariance matrix, and R^d is the domain of data.

$$v^* = \arg \max_{y \in R^d} v^T M v. \quad (5)$$

Each vector that is found by equation 4 will be treated as the principal components (PCs), the first component will dominate the most significant percentage of data variance, and the following components will have the decreasing data variance for the whole signal, the PCs could be found by the singular value decomposition, and all of it are the linear combinations of the initial signal but not unrelated to each other. This method has the advantage of decreasing the disturbance from noise or small signals, and it is often utilized with a multidimensional reduction which retains the most important components before the following procedure. This method can eliminate the multi-collinearity because the PCs are uncorrelated. However, the PCR has the drawback that the PCs are obtained through only variables themselves, and it cannot make sure that the variables correlate with the observation data, which often makes lower precision outcomes from the fitting model. Consequently, it is trendy that combining with other discriminant methods for target detecting or classifying.

PLSR¹⁹ was proposed to overcome the drawback of PCR; PLSR extracts the dependent variables by confirming their predictive abilities. These orthogonal factors, which are also known as latent variables (LVs), are obtained according to the correlation with the predicting variable. The general basic model of PLSR can be denoted as equation 6 and 7, where X is the matrix of predictors, Y is a matrix of response, T and U are the projection of X and Y , P and Q are coefficient matrices, E and F are the error terms. The main idea of the PLSR is to look for the proper projection that has the maximum correlation shown in equation 8. Different from the PCA, which extracts the PCs with a high covariance variable, the PLSR is more feasible in the condition of multi-collinearity between the variables. Thus, the number of LVs is usually less than PCs in PCR regression.

$$X = TP^T + E \quad (6)$$

$$Y = UQ^T + F \quad (7)$$

$$t, u = \arg \max_{T, U} Cov(T, U). \quad (8)$$

Feedforward Neural Network

The concept of the feedforward neural network (FNN) model was introduced by Rosenblatt in 1958²⁰. In this work, FNN was applied to analyze $N \times W$ data where N denotes the number of samples and W denotes the number of bands. The proposed FNN model, illustrated in the Figure 6, was designed with the input layer consisting of individual hyperspectral bands of input data, hidden layers containing 2048 and 512 neurons with the batch normalization layer and dropout layer to prevent overfitting in model training, and one neuron for the output layer, each layer activated with the Rectified Linear Unit, ReLU(α), activation function. In our study, there were still some outliers in our dataset. Therefore, the Root Mean Squared Log Error (RMSLE) was used to evaluate the error during training. Referring to Root Mean Squared Error (RMSE), Root Mean Squared Logarithmic Error (RMSLE) is yielded by

$$RMSLE = \sqrt{\frac{1}{n} \sum_i^n (\log(X_i + 1) - \log(Y_i + 1))^2} \quad (9)$$

where X is the predicted value, and Y is the target value. This measure shows the ability to nullify effects of outliers by reducing the difference between and with the natural logarithm.

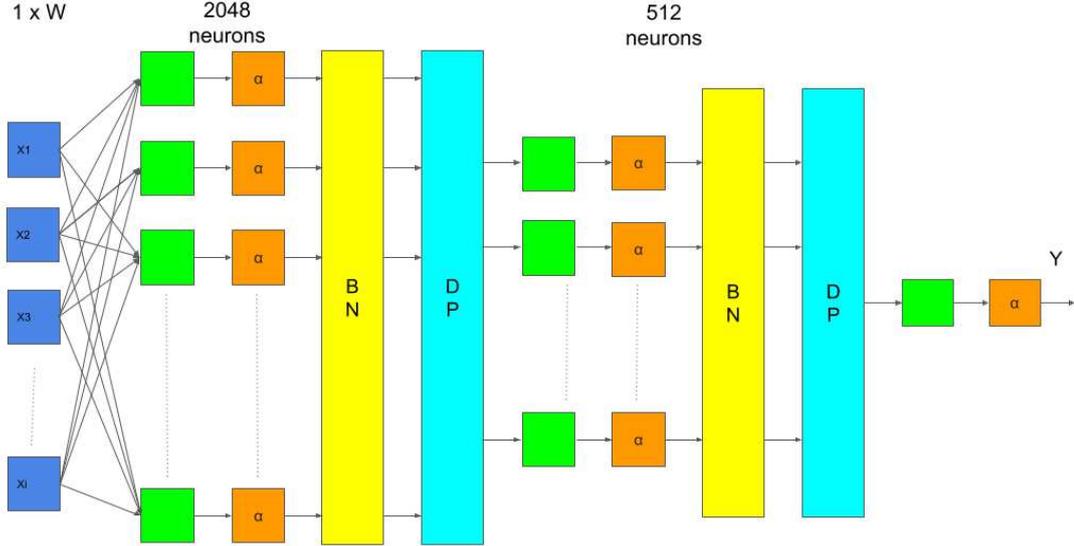


Figure 6. The figure shows the architecture of the proposed feedforward neural network for sugariness. Input is a 1-D spectra data with size for the number of bands. Hidden layers contain one layer with 2048 neurons and another layer with 512 neurons, each layer connected with a ReLU activation function(α), batch normalization layer (BN), and dropout layer (DP). The final output Y is the sugariness prediction result of the model.

Convolutional Neural Networks

Convolutional Neural Network (CNN) was also one of the approaches applied to our dataset. The modern CNN architecture was proposed by LeCun in 1998²¹. Its strong ability to extract spatial features makes it achieve state-of-the-art performance in computer vision tasks. In this study, however, instead of extracting spatial features, the 2-D CNN model was used to extract spectral features from 3-D HSI data with size (W, L, K) , where W denotes the width value with constant value 20, L denotes the length of bands and K denotes the number of channels of a hyperspectral image. Note that W and K are actually the width and length of the original hyperspectral image, but the model was designed to consider the bands as the length and width as the input data channels. The entire architecture is shown in the Figure 7. A model was designed to extract spectral features from the hyperspectral image, which had four convolutional layers with filter size 1×2 and two Maxpooling layers with size 1×2 and fully connected layers with 1×1 dense layer for output, all the convolutional layers with Batch Normalization (BN) and Rectified Linear Unit (ReLU) as the non-linear activation function. Noted that the filters with size 1×2 of each convolutional layer were used to slide along the space (W, L) , and Maxpooling layers calculated the maximum from each 1×2 patch on space (W, L) in order to extract spectral features instead of spatial features of the HIS. Moreover, the output feature map obtained from the last Maxpooling layer then flattened to a 1-D array and fed into fully-connected layers. Finally, the final activation layer was also a ReLU function, it gave a Brix value of the wax apple sample, and the error is evaluated with Root Mean Squared Logarithmic Error (RMSLE) in equation 9.

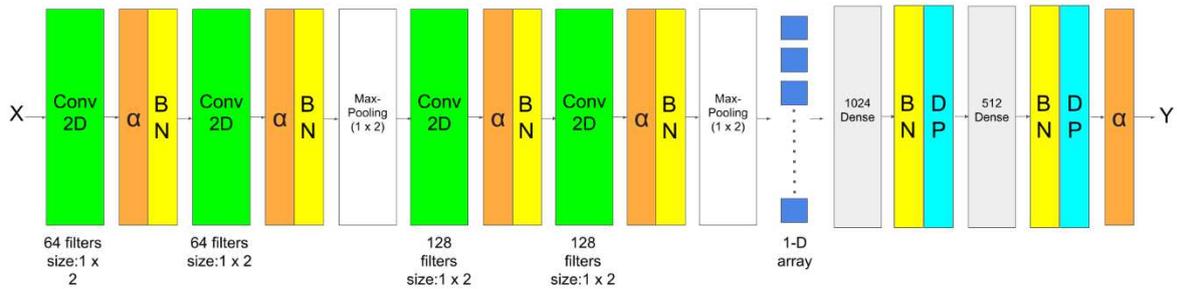


Figure 7. The figure shows the architecture of the proposed convolutional network for sugariness prediction. Input X is a 3-D HSI sample, and the network contains four convolutional layers with a ReLU activation function. Batch normalization (BN) layers and Dropout (DP) layers are used to prevent overfitting. In the last Maxpooling layer, its output is a 1-D array obtained by calculating the maximum from each 1×2 patch of the feature map then flattening. The 1-D array then was fed to fully-connected layers with two dense layers with BN and DP following to regress the sugariness prediction result Y .

Results

Evaluation by Data Visualization

In the experiments, *t*-SNE was applied to reduce the dimensionality of HSIs and visualize the distribution of the datasets. The datasets used to make dimension reduction were the 1-D array datasets averaged from 3-D HSIs, because it would be time-consuming, and the results were not too much different with 3-D HSIs. The entire dataset contained 932 one-dimensional reflectance spectral data arrays for different bands. Before starting the *t*-SNE algorithm in the experiment, the Principal Component Analysis (PCA) was used to initialize the value for dimension reduction results. Then the *t*-SNE models were trained with the perplexity value 20, the 30000 iteration times, and the learning rate 1000. Figure 8 shows the visualization results of four different datasets.

Figure 8(a), (b), and (c) show the *t*-SNE results of datasets with wavelength 400-1700nm, 400-1000nm, and 900-1700nm, although the data points of sugariness above 14 °Brix and under 10 °Brix seem to be clustered, the clusters of data points with sugar content between 14 °Brix and 10 °Brix seem to be ambiguous. On the other hand, the clusters of *t*-SNE results shown in the Figure 8(d) are much more ambiguous than the different datasets. Therefore, it would indicate that the spectra data of wavelength 400-1000nm were more significant than the spectra data of wavelength 900-1700nm to regress the sugariness value. This *t*-SNE result reveals that the spectra data of visible spectroscopy (VIS) bands could contain more significant features toward predicting sugariness than the near-infrared (NIR) spectroscopy bands.

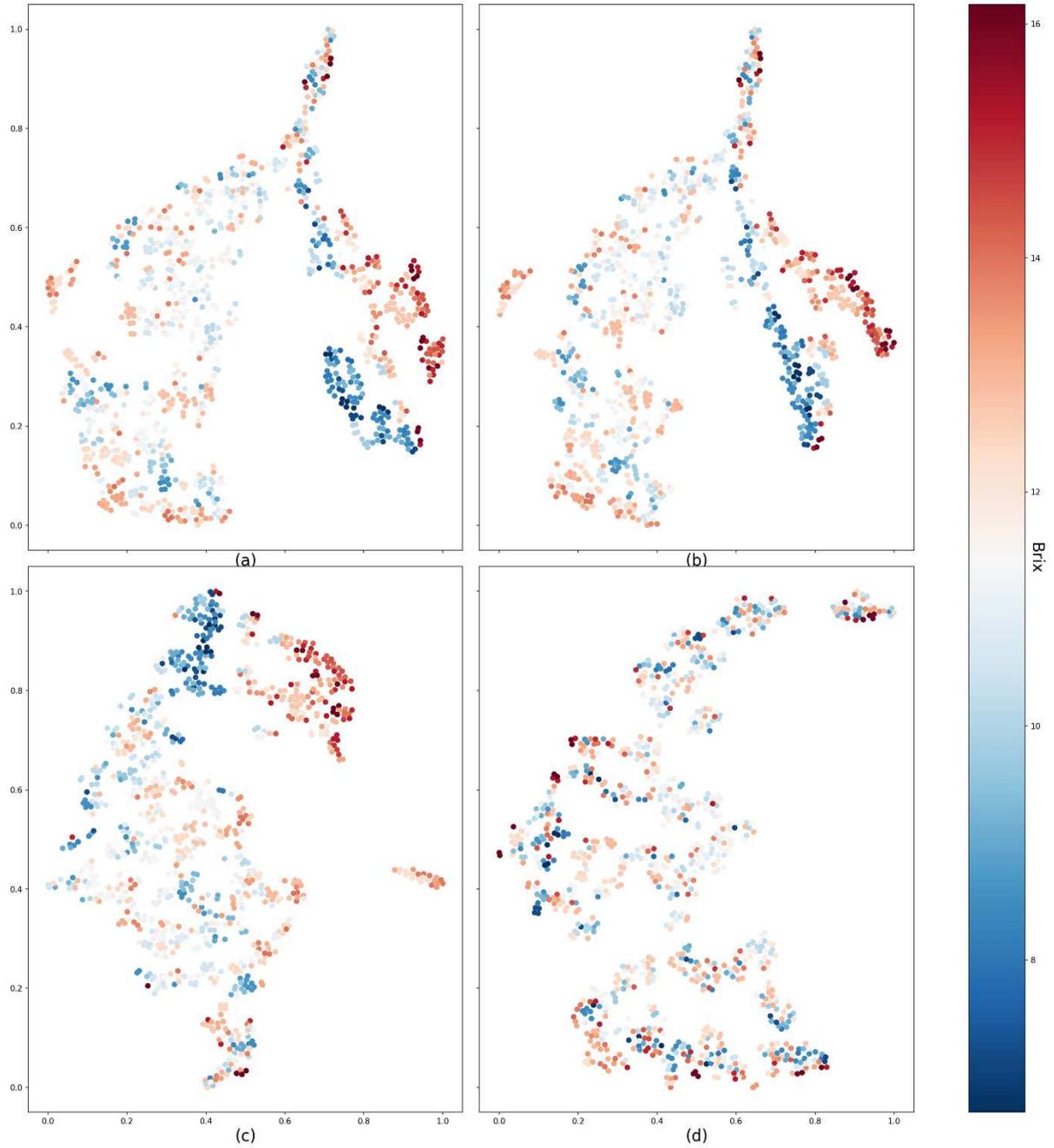


Figure 8. The figure shows the results of t -SNE on the four datasets, (a) to (d) are the datasets of wavelength range 400-1700nm, 400-1000nm, 400-700nm, and 900-1700nm, the color bar in the right side denoted the Brix value of each data point. All the t -SNE results are obtained with the perplexity value 20, the 30000 iteration times, and the learning rate 1000.

Evaluation over Different Models

The weights and kernels in FNN and CNN models were initialized with Glorot initialization²² and regularized with L2 regularization²³ with regularization parameter 3×10^{-5} while training. Adam optimizer²⁴ was used with mini-batch 128 and 3000 epochs for all the models to optimize the weights of each layer of FNN models and kernels of CNN models. Learning rate decay method applied in the experiments for optimizing models, the initial learning rate value for first epoch was 1×10^{-3} , and an exponential decay $lr_t = lr_{t-1} * e^{(-k*t)}$ (lr denotes as the learning rate, t denotes as the epoch number, k denoted as a constant of 1×10^{-4}) was applied for following epochs.

The bar chart Figure 9 shows the RMSLE of each method on the validation set. For testing each of the deep learning models, the trained weights were chosen at the step with the smallest validation loss in the

optimization process. The minimum of the validation loss of each FNN model for 400-1700nm, 400-1000nm, 400-700nm, and 900-1700nm were 0.005463 at step 2734, 0.005917 at step 2736, 0.005707 at step 2540, and 0.007042 at step 2980. The model for 400-1700nm had the lowest loss, and the model for 900-1700nm had the highest loss, so it was expectable that the FNN model for 400-1700nm would have better performance on the test dataset than the other FNN models. There were also four models for CNN. The minimum of the validation loss (RMSLE) of the 400-1700nm model was 0.005165 at step 2844, 0.005478 at step 2732 for the model of 400-1000nm, 0.006077 at step 2092 for the model of 400-700nm, and 0.015015 at the step 2994 for the model of 900-1700nm. As same as the results of FN models, the CNN model for 400-1700nm had the lowest loss, and the model for 900-1700nm had the highest loss. Comparing the results of FN with the results of CNN models, the CNN models would have better performance than FNN models. All the deep learning modeling was implemented using the Keras²⁵ and Tensorflow backend²⁶ on an NVIDIA RTX 3090 GPU.

PCR and PLSR were used to make a comparison with deep learning methods on the same training datasets. The PCR result showed that the validation loss (RMSLE) on four validation datasets was 0.165022 for the 400-1700nm model, 0.164529 for the 400-1000nm model, 0.156733 for the 400-700nm model, and 0.166839 for the 900-1700nm model. On the other hand, validation losses of PLSR were 0.153204, 0.152438, 0.150596, and 0.16576 on four different datasets. These results indicate that the PLSR method would have better performance than PCR on the collected datasets, and deep learning methods would be much better than PCR and PLSR.

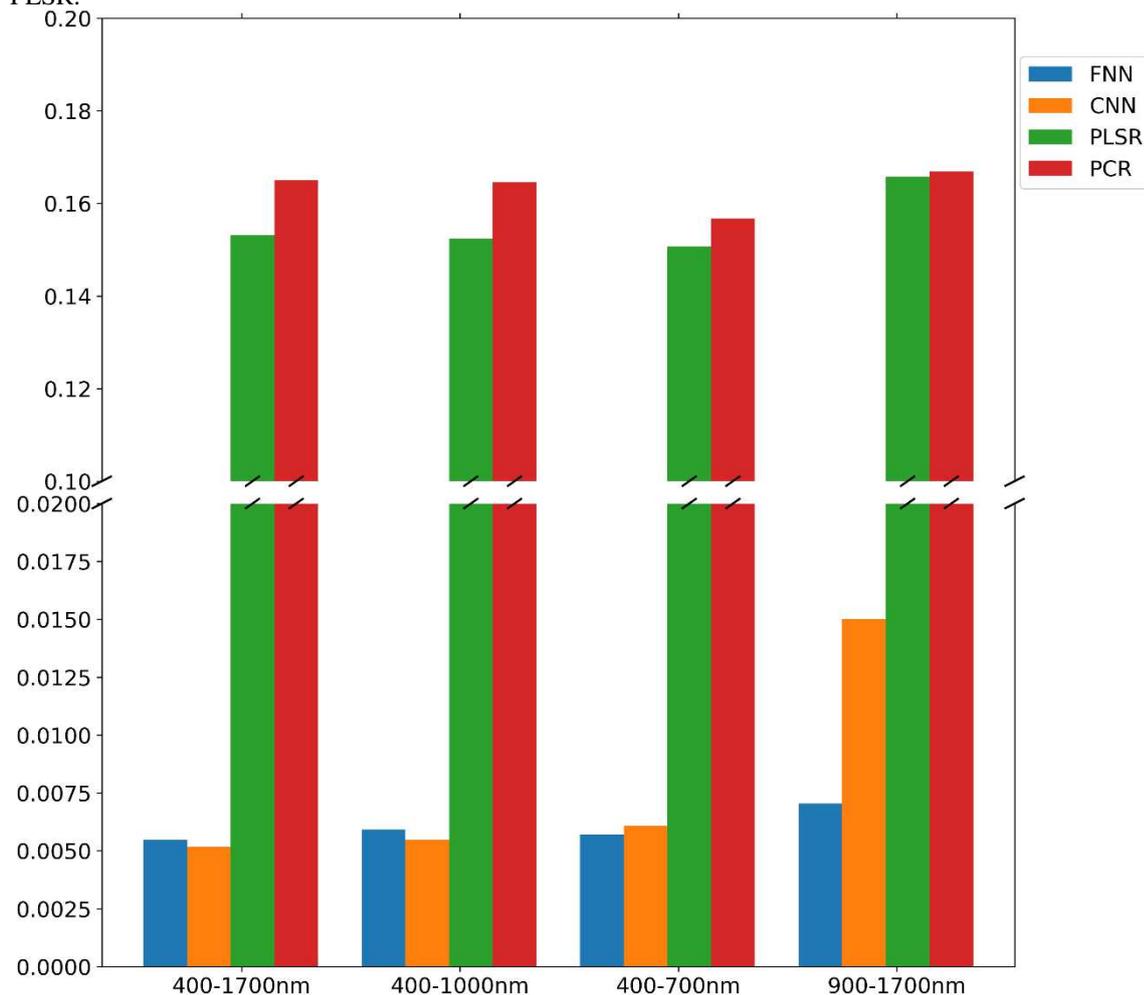


Figure 9. The bar chart shows the validation loss (RMSLE) of each proposed model. In the optimization process, the weights with the smallest validation loss are recorded as the best sets. This validation error presented in the chart indicates that FNN and CNN would have superior performance compared with PCR and PLSR on the datasets.

The test results of Brix of PCR, PLSR, FNN, and CNN models in each interval shown in Table 2. First of all, although PCR and PLSR showed their good capability in predicting sugar content in Brix interval 11-12, the test errors in other Brix intervals and averaged test errors were not competitive with errors of FNN and CNN models. This phenomenon indicated that the PCR and PLSR could have an overfitting problem on this dataset and

improperly detect sugar content of wax apple. The FNN models had the average error with ± 0.597 °Brix on the dataset 400-1700nm, 400-1000nm with ± 0.610 °Brix, 400-700nm with ± 0.623 °Brix, and 900-1700nm with ± 0.739 °Brix. The test results showed that FNN models trained using the dataset of wavelength range 400-1700nm had adequate performance, and wavelength range 900-1700nm would not correlate with sugar content prediction using FNN. The test results of CNN models were also similar to the test results of FNN models. The test error was ± 0.552 °Brix on the dataset of wavelength 400-1700nm, ± 0.616 °Brix on the dataset of wavelength 400-1000nm, ± 0.587 °Brix on the dataset of wavelength 400-700nm, and ± 0.7396 °Brix on a dataset of wavelength 900-1700nm. The test results of CNN models were superior to FNN models, and both of the results of FNN and CNN showed that spectra data corresponding to 400-1700nm, 400-1000nm, and 400-700 had better correlation than data corresponding to 900-1700nm. This conclusion indicated that VIS bands with wavelength 400-700nm might be more correlative to predict sugar content than NIR bands with wavelength 900-1700nm. This phenomenon is also shown in validation results and accuracy of k-NN classification on *t*-SNE results. On the other hand, the minimal error of PCR and PLSR were ± 1.444 °Brix and ± 1.379 °Brix on the dataset of wavelength 400-700nm, but both of them had their best (± 0.180 °Brix and ± 0.218 °Brix) in Brix interval 11-12 on 900-1700nm. Therefore, PCR and PLSR were more severe with overfitting problems on predicting sugar content using spectra data with wavelength range 900-1700nm than the other band range. The averaged test error of each learning method on different band ranges was presented in the Figure 10. The outcomes of test data indicated that the deep learning methods FNN and CNN were more remarkable than PCR and PLSR on predicting sugariness, and these results were also reflected in validation results.

Table 2. The table shows all the test errors in different band ranges and Brix intervals. The average test results of FNN and CNN indicate that the models have a decent complexity for data sets. On the other hand, PCR and PLSR would have underfitting problems on the datasets.

Method	Band Range	Sugar Content in each Brix Intervals(\pm °Brix)					
		0-10 43 Samples	10-11 32 Samples	11-12 43 Samples	12-13 43 Samples	13- 43 Samples	ALL 162 Samples
PCR	400-1700nm	2.727	1.061	0.218	1.103	2.430	1.532
	400-1000nm	2.678	1.051	0.233	1.107	2.441	1.526
	400-700nm	2.504	0.942	0.339	1.018	2.287	1.444
	900-1700nm	2.797	1.063	0.180	1.080	2.394	1.526
PLSR	400-1700nm	2.449	0.927	0.360	0.979	2.119	1.391
	400-1000nm	2.410	0.954	0.371	0.965	2.098	1.382
	400-700nm	2.405	0.958	0.375	0.218	2.369	1.379
	900-1700nm	2.821	1.049	0.218	1.100	2.369	1.536
FNN	400-1700nm	0.642	0.646	0.511	0.564	0.637	0.597
	400-1000nm	0.625	0.633	0.437	0.474	0.889	0.610
	400-700nm	0.732	0.693	0.355	0.463	0.892	0.623
	900-1700nm	0.710	0.656	0.576	0.701	1.032	0.732
CNN	400-1700nm	0.609	0.474	0.412	0.421	0.660	0.552
	400-1000nm	0.627	0.487	0.451	0.516	0.965	0.616
	400-700nm	0.551	0.408	0.462	0.555	0.914	0.587
	900-1700nm	1.519	0.725	0.496	0.480	1.543	0.965

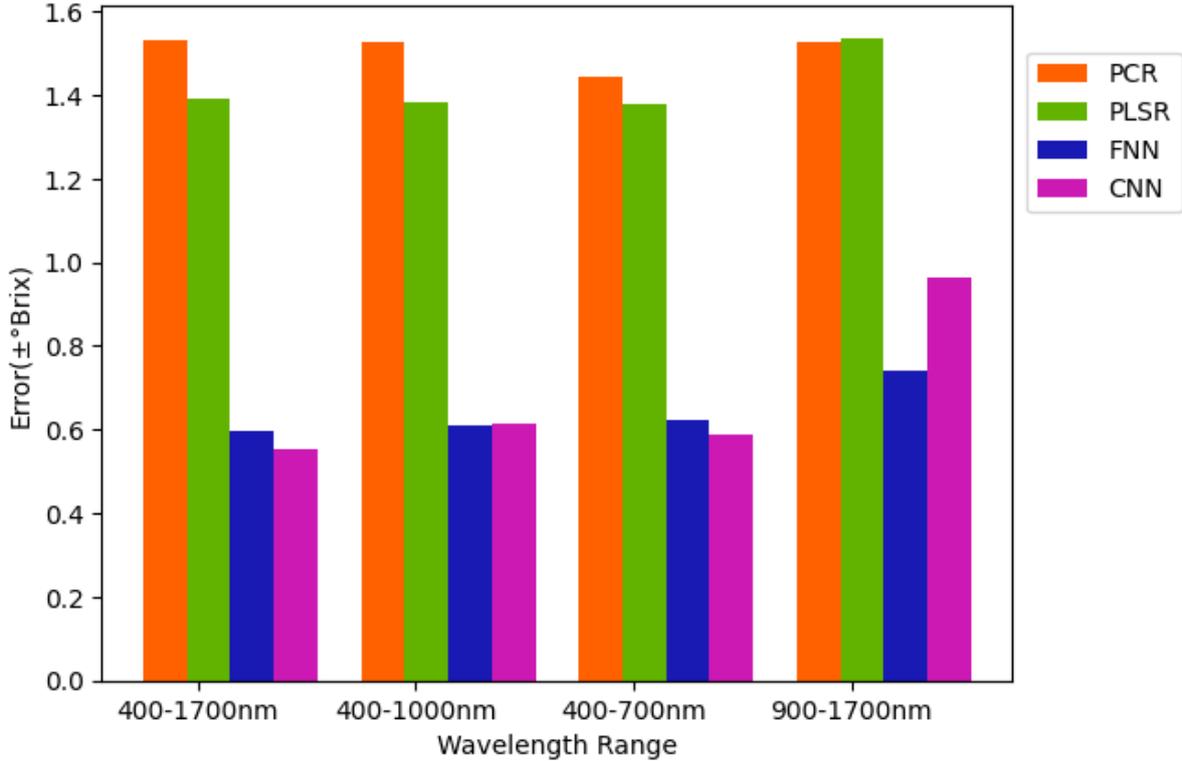


Figure 10. The bar chart shows that the test result of each model. The test results of FNN and CNN have more minor errors than PCR and PLSR. It indicates that the performance of non-linear deep learning models is better than linear models on the data sets.

Evaluation of the Learning Results by Visualizing the Outputs before the Last Layer

Visualization was also a decent way to evaluate the proposed deep learning models. Therefore, *t*-SNE was applied to visualize the outputs of the layer before the last regression output layer. As same as the previous implementation of *t*-SNE, the PCA was used to be the initial value for dimension reduction results, and the perplexity value was set 20, the 30000 iteration times, and the learning rate 1000 were used in this implementation. Figure 11 and 12 show the final visualization results of MLP and CNN models corresponding to four different wavelength ranges.

For MLP models, the output size of the layer before the last output layer was 512. Therefore, the output was an array with size 1×512 . Then all the outputs from feeding all the data sets ($N \times 512$) were collected and scaled to N data points visualized in the Figure 11. In the Figure 11(a), (b), and (c) show the final evaluation results corresponding to wavelength range 400-1700nm, 400-1000nm, and 400-700nm and although there are still some vague clusters, each of the results presents the proper color gradation generally from deep red to deep blue with respect to high sugar content value to low sugar content value, which represents the acceptable learning ability of MLP models on those datasets. Although the visualization results were better than the *t*-SNE results of the original data, ambiguous clusters are still contained in the Figure 11(d). Therefore, this result elaborates that wavelength 900-1700nm spectra data would not be advantageous for sugar content prediction.

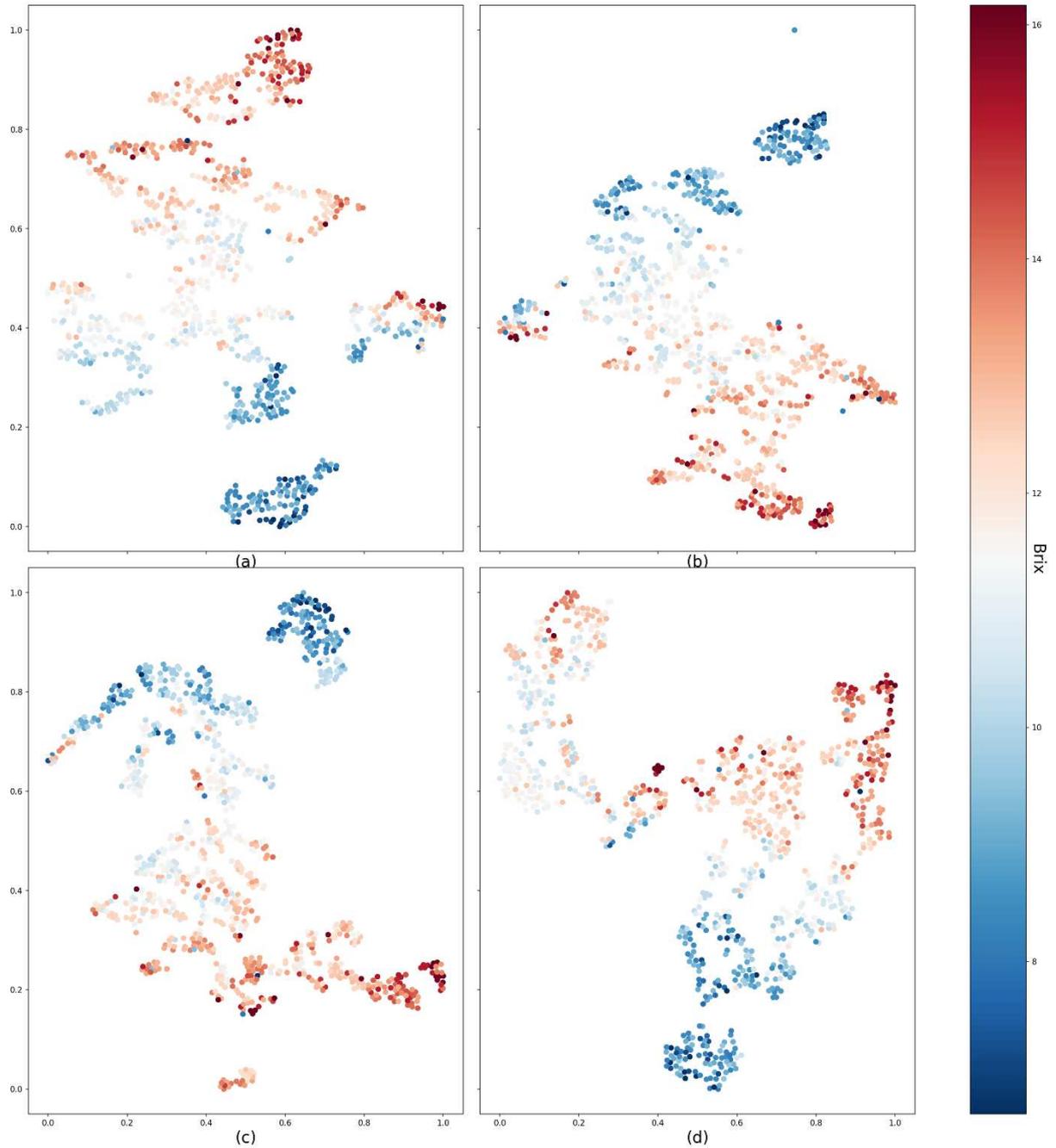


Figure 11. The figure shows the result of t -SNE on the outputs obtained by feeding the four datasets to FNN models, (a) to (d) are the datasets of wavelength range 400-1700nm, 400-1000nm, 400-700nm, and 900-1700nm. The color bar on the right side is denoted as the Brix value of each data point. Compared to t -SNE results on original datasets, the figures(a) to (c) generally show the fine color gradation from deep red to deep blue, relatively. Relatively figure(d) contains more vague clusters than others.

On the other hand, Figure 12(a) to (d) visualize the t -SNE results on the outputs of the layer before the last output layer of CNN models corresponding to the wavelength range 400-1700nm, 400-1000nm, 400-700nm, and 900-1700nm. According to the CNN architecture shown in the Figure 7, each of the data points in the Figure 12 was also obtained by scaling output from 1×512 to 1×2 using t -SNE. The t -SNE visualization results gave the fine gradational effect that is shown in the Figure 12(a) to (d). However, more outliers appeared in clusters of t -SNE results in the Figure 12(d). These results showed that the CNN models were capable of learning features toward the sugar content prediction by showing better clustering results on outputs before the last regression output layer than the ones on original data.

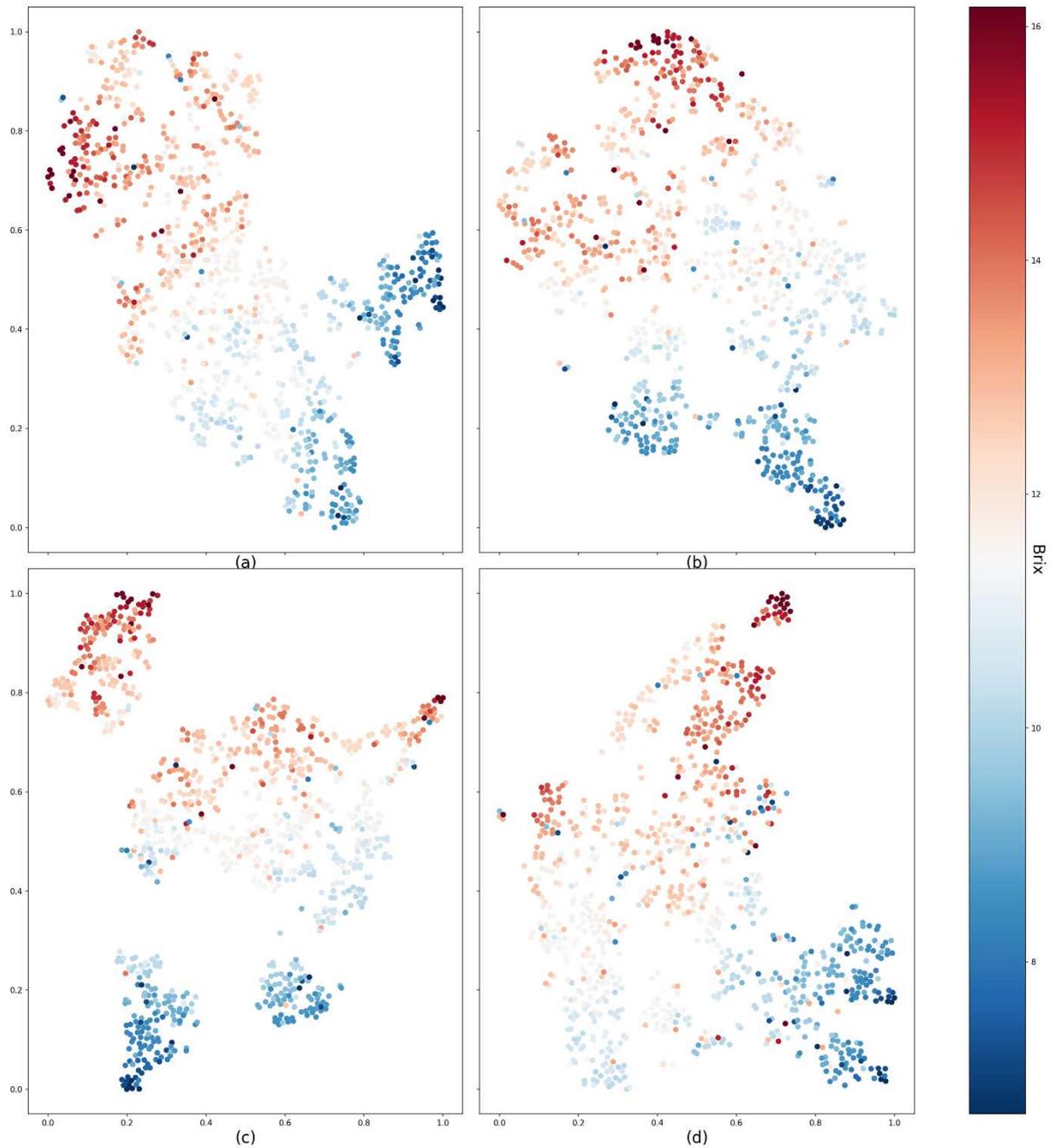


Figure 12. The figure shows the t -SNE results on the outputs obtained by feeding the four datasets to CNN models are visualized in (a) to (d) corresponding to wavelength range 400-1700nm, 400-1000nm, 400-700nm, and 900-1700nm. The color bar on the right side is represented as the Brix value of each data point. Compared to t -SNE results on original datasets, the figure(a) to (d) generally show the better color gradation from deep red to deep blue. Although there are some outliers contained in clusters in figure(d), the t -SNE visualization results in the figure(a) to (d) are much better than ones in the Figure 8(a) to (d).

Conclusions and Discussion

This study aims to predict the sugariness of *Syzygium samarangense* from hyperspectral images using deep learning methods. The HSIs were collected from the bottom part of the wax apple and adequately processed to 1-D spectra datasets and 3-D HSIs datasets. The spectra data were adequately sampled and divided into 400-1700nm, 400-1000nm, 400-700nm, and 900-1700nm to facilitate learning from the corrected datasets, then t -SNE was used to evaluate the performance of the spectral band. In the modeling stage, the FNN and CNN models were designed for analyzing the HSIs datasets. For comparing with deep learning methods, PCR and PLSR were used on the same datasets. The best results of FNN and CNN were 0.552 and 0.597 (\pm °Brix) on datasets with 400-1700nm, relatively, PCR and PLSR had their best result on datasets with 400-700nm, which were 1.379 and 1.444

(\pm °Brix). In this study, FNN and CNN models outperformed PCR and PLSR models. Otherwise, all their well-performed results were obtained with the spectra data with VIS bands contained, and the results with only using NIR bands(900-1700nm) were not that acceptable. Therefore, this study revealed that VIS bands would be more critical in predicting sugar content and showed the potential performance on acquiring sugar content of wax apple with spectra data corresponding to VIS bands using a deep learning model. Moreover, the improved *t*-SNE results on the final regression outputs before the last layer of proposed deep learning models show that FNN and CNN models would adequately learn how to predict sugariness from the HSI datasets.

Compared with the previous study¹⁵, this paper provided the average error and the error in each Brix interval within one °Brix. Although their works had good prediction results with total SEV=0.381 and 0.426 °Brix for ICA and PLSR using only NIR bands, this study indicated that the error would significantly different within different Brix intervals. Therefore, this study which provided errors of different Brix intervals would be more detailed than theirs.

In future work, it will be significant to select proper bands for fabricating a portable device to detect sugariness. The light source with 400-700nm and the CNN model would be a good start for constructing a mobile device because the light source with 400-700nm is easily available. Moreover, this combination shows adequate performance on prediction sugariness of wax apple in this study with the minor error in Brix interval 0-10°Brix and 10-11°Brix are \pm 0.551 and \pm 0.408. These results indicate that the CNN model using 400-700nm spectra data would have the capability to predict if sugariness is below 10°Brix or not, which would be similar to the human tongue.

Acknowledgment This work is supported by the Council of Agriculture, Taiwan under Project (COA 108AS-13.2.11-ST-a6, COA 107AS-14.2.11-ST-a3) and by the Ministry of Science and Technology, Taiwan under Project (MOST 109-2321-B-009-008 -s, MOST 108-2321-B-009-001 -, MOST 108-2221-E-009-125 -, MOST 107-2321-B-009-002, MOST 107-2622-E-009-015-CC2, MOST 107-2221-E-009-155), and National Yang Ming Chiao Tung University.

Conflict of interest All authors declare that they have no conflict of interest to declare.

Declarations

- a. The collection, research, and studies of the cultivated plants of this study complied with the relevant institutional, national, and international guidelines and legislation.
- b. The farmers permitted the transaction of plants in studying usage.
- c. If request, the data sets of HSI of wax apples will be available.

References

1. Taghizadeh, M., Gowen, A. A., & O'Donnell, C. P. Comparison of hyperspectral imaging with conventional RGB imaging for quality evaluation of Agaricus bisporus mushrooms. *Biosystems engineering*, 108(2), 191-194. <https://doi.org/10.1016/j.biosystemseng.2010.10.005> (2011).
2. Bioucas-Dias, J. M., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., & Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geoscience and remote sensing magazine*, 1(2), 6-36. <https://doi.org/10.1109/MGRS.2013.2244672> (2013).
3. Akhtar, N., & Mian, A. Nonparametric coupled Bayesian dictionary and classifier learning for hyperspectral classification. *IEEE transactions on neural networks and learning systems*, 29(9), 4038-4050. <https://doi.org/10.1109/TNNLS.2017.2742528> (2017).
4. Zhong, P., & Wang, R. Jointly learning the hybrid CRF and MLR model for simultaneous denoising and classification of hyperspectral imagery. *IEEE Transactions on Neural Networks and Learning Systems*, 25(7), 1319-1334. <https://doi.org/10.1109/TNNLS.2013.2293061> (2014).
5. Peiris, K. H. S., Dull, G. G., Leffler, R. G., & Kays, S. J. Near-infrared spectrometric method for nondestructive determination of soluble solids content of peaches. *Journal of the American Society for Horticultural Science*, 123(5), 898-905. <https://doi.org/10.21273/JASHS.123.5.898> (1998).

6. Mendoza, F., Lu, R., Ariana, D., Cen, H., & Bailey, B. Integrated spectral and image analysis of hyperspectral scattering data for prediction of apple fruit firmness and soluble solids content. *Postharvest Biology and Technology*, 62(2), 149-160. <https://doi.org/10.1016/j.postharvbio.2011.05.009> (2011).
7. Park, B., Abbott, J. A., Lee, K. J., Choi, C. H., & Choi, K. H. Near-infrared diffuse reflectance for quantitative and qualitative measurement of soluble solids and firmness of Delicious and Gala apples. *Transactions of the ASAE*, 46(6), 1721. <http://dx.doi.org/10.13031/2013.15628> (2003).
8. Kempes, B., Leon, L., Best, S., De Baerdemaeker, J., & De Ketelaere, B. Assessment of the quality parameters in grapes using VIS/NIR spectroscopy. *Biosystems Engineering*, 105(4), 507-513. <https://doi.org/10.1016/j.biosystemseng.2010.02.002> (2010).
9. 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. *Artificial Intelligence in Agriculture*, 4, 31-38. <http://dx.doi.org/10.1016/j.aiia.2020.04.003> (2020).
10. Itakura, K., Saito, Y., Suzuki, T., Kondo, N., & Hosoi, F. Estimation of citrus maturity with fluorescence spectroscopy using deep learning. *Horticulturae*, 5(1), 2. <https://doi.org/10.3390/horticulturae5010002> (2019).
11. Viegas, T. R., Mata, A. L., Duarte, M. M., & Lima, K. M. Determination of quality attributes in wax jambu fruit using NIRS and PLS. *Food Chemistry*, 190, 1-4. <https://doi.org/10.1016/j.foodchem.2015.05.063> (2016).
12. Tu, S., Pang, J., Liu, H., Zhuang, N., Chen, Y., Zheng, C., Wan, H., & Xue, Y. Passion fruit detection and counting based on multiple scale faster R-CNN using RGB-D images. *Precision Agriculture*, 21(5), 1072-1091. <https://doi.org/10.1007/s11119-020-09709-3> (2020).
13. Fajardo, M. & Whelan, B.M. Within-farm wheat yield forecasting incorporating off-farm information. *Precision Agriculture* 22, 569–585 <https://doi.org/10.1007/s11119-020-09779-3> (2021).
14. Marani, R., Milella, A., Petitti, A., & Reina, G. Deep neural networks for grape bunch segmentation in natural images from a consumer-grade camera. *Precision Agriculture*, 22(2), 387-413. <https://doi.org/10.1007/s11119-020-09736-0> (2021).
15. Chuang, Y. K., Chen S., Lo, Y. M., Tsai, C. Y., Yang, I. C., Chen, Y. L., Pan, P. J. & Chen, C. C. Integration of independent component analysis with near infrared spectroscopy for rapid quantification of sugar content in wax jambu (*Syzygium samarangense* Merrill & Perry). *J Food Drug Anal*, 20(855), e64. <http://dx.doi.org/10.6227%2fjfd.2012200415> (2012).
16. Li, Y. S., Yan, Y.J., Shih, R. S., Chang, C. H., Chen, T. C., Chen, Y. C., Huang, C. C., Lin, S.G., and Ou-Yang, M. Development and verification of the coaxial heterogeneous hyperspectral system for the Wax Apple tree, *International Instrumentation and Measurement Technology Conference*, 1-5. <https://doi.org/10.1109/I2MTC.2019.8826836> (2019).
17. Savitzky, A., Golay, M. J. E. Smoothing and differentiation of data by simplified least squares procedures, *Analytical Chemistry*, 36(8), 1627-1639. <https://doi.org/10.1021/ac60214a047> (1964).
18. Van der Maaten, L., & Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research*, 9(11). <http://www.jmlr.org/papers/v9/vandermaaten08a.html> (2008).
19. Wold, S., Sjöström, M., & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109-130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1) (2001).
20. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain, *Psychological Review*, 65, 386-408. <http://dx.doi.org/10.1037/h0042519> (1958).
21. LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791> (1998).

22. Glorot, X., & Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proc. of International Conference on Artificial Intelligence and Statistics*, 249-256. <http://proceedings.mlr.press/v9/glorot10a.html> (2010).
23. Cortes, C., Mohri, M., & Rostamizadeh, A. L2 regularization for learning kernels. *arXiv preprint arXiv:1205.2653*. <https://arxiv.org/abs/1205.2653> (2012).
24. Kingma, D. P., & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. <https://arxiv.org/abs/1412.6980> (2014).
25. Gulli, A., & Pal, S. *Deep learning with Keras*. Packt Publishing Ltd. <https://keras.io/> (2017).
26. TensorFlow. API TensorFlow Core v2.3.0 - Python. Retrieved March 01, 2020, from https://www.tensorflow.org/api_docs/python/tf. (2020).