 doi:10.20944/preprints202008.0378.v1

Article

Nondestructive detection for egg freshness based on hyperspectral scattering image combined with ensemble learning

DAI Dejian[1,2], JIANG Tao[1], LU Wei[1,2*], SHEN Xuan[1], XIU Rui[1], ZHANG Jingwei[1]

[1] College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210031, China; djdai@njau.edu.cn (D.D.); 32217208@njau.edu.cn (J.T.); 32217320@njau.edu.cn (S.X.); 32217325@njau.edu.cn (X.R.); 1113913031@qq.com (Z.J.)

[2] Jiangsu Province Engineering Laboratory of Modern Facility Agriculture Technology and Equipment, Nanjing 210031，China

* Correspondence: njaurobot@njau.edu.cn

Abstract: Scattering hyperspectral technology is a nondestructive testing method with many advantages**.** Here, we propose a method to improve the accuracy of egg freshness, research the influence of incident angles of light source on the accuracy and explain its mechanism. A variety of weak classifiers classify eggs based on the spectra after preprocessing and feature wavelength extraction to obtain three classifiers with the highest accuracy. The three classifiers are used as metamodels of stacking ensemble learning to improve the highest accuracy from 96.25% to 100%. Moreover, the highest accuracy of scattering, reflection, transmission and mixed hyperspectral of eggs are 100.00%, 88.75%, 95.00% and 96.25%, respectively, indicating that the scattering hyperspectral for egg freshness detection is better than that of the others. In addition, the accuracy is inversely proportional to the angle of incidence due that the smaller the incident angle, the camera collects a larger proportion of scattering light, which contains more biochemical parameters of an egg than that of reflection and transmission. These results are very important for improving the accuracy of non-destructive testing and selecting the incident angle of the light source, and have potential applications in online non-destructive testing.

Keywords: Egg freshness; Hyperspectral detection; Hyperspectral scattering imaging； Ensemble learning

1. Introduction

The freshness of eggs is related to their nutritional value. It is the most concerned index of processing companies and consumers, and an important index in transportation and processing [1]. It can be detected by traditional biochemical methods, but they are destructive, time-consuming and inefficient. Therefore, nondestructive testing technology has significant advantages in the detection of egg freshness and has attracted wide attention. Currently, egg freshness is tested by nondestructive techniques of spectral analysis [2-3], dielectric property [4-5], electronic nose [6-7], machine vision [8-9] and hyperspectral testing [10-12]. Specially, machine vision method was established for egg freshness with R (correlation coefficient) of 0.8653 (Sun, L.; et al., 2014) [8]. The prediction model was established by near infrared spectroscopy with R of 0.879 (Lin, H.; et al., 2012) [13]. The freshness model was established by testing the volatile concentration of eggs by electronic nose with a low efficiency, thus it is not suitable for the dynamic testing of production line (Yimenu, S.; et al, 2017) [7]. The egg freshness was tested by reflectance near infrared hyperspectral with R of 0.879 (Suktanarak; S. et al, 2017) [10], which could achieve rapid and nondestructive classification of egg freshness. However, the model precision could not be further improved due to the great influence of eggshell colors. Therefore, hyperspectral technology can effectively improve the test accuracy by map fusion and information dimension expansion, but different light sources have a great influence on the measurement results [14]. The spectral scattering imaging by optical fiber is used to study the internal light propagation paths in apples and tomatoes to realize nondestructive testing of the surface and internal defects of fruits and vegetables (Renfu L.; et al. 2017) [15].

Herein, we proposed a method to improve the accuracy of egg freshness based on hyperspectral scattering imaging, researched the influence of incident angles on the accuracy and explained its mechanism. Have found that stacking ensemble learning could be used to improve the highest accuracy of egg freshness, and the accuracy is inversely proportional to the incident angle. These are useful for improving the accuracy of a classifier, important for selecting the incident angle of a light source with an accuracy, and have potential applications in online nondestructive testing.

2. Materials and Methods

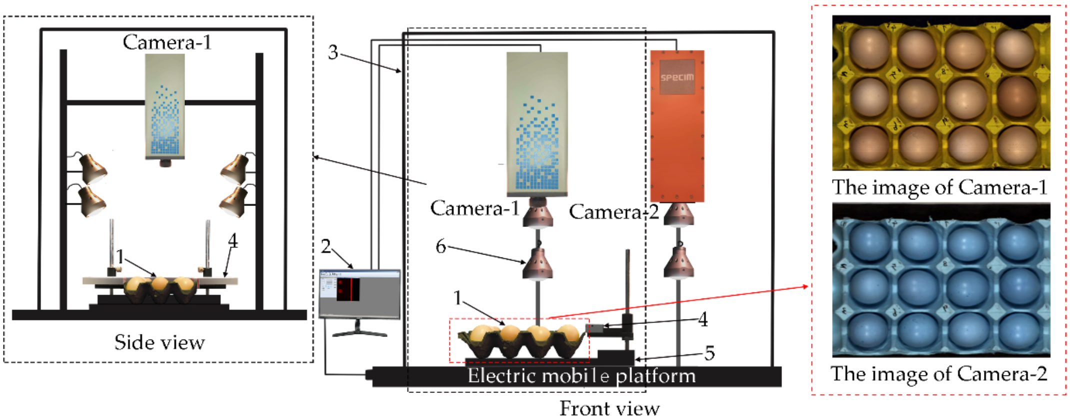
2.1. Experimental materials

350 eggs (pink shell, mass: 31.5–46.6 g, equatorial diameter: 32.8–41.9 mm) were purchased from Panchu Mechanized Chicken Farm, Nanjing, Jiangsu Province, China. They were all produced on the day of purchase, and stored at room temperature after cleaning. These eggs were divided into two groups, the data and the calibration group with 200 and 150 eggs, respectively. The data group was used to collect hyperspectral images, and the calibration group was used to measure the Haugh unit.

2.2. Hyperspectral imaging system

The hyperspectral instrument is GaiaSorter-Dual "Gaia" dual-camera all-band hyperspectral sorter. Its main components include a uniform light source, a dual spectrum camera, an electronic control transfer module, a computer with a control software, etc. The dual spectrum Camera include two hyperspectral cameras, the Camera 1 (Image-λ-V10E, wavelength range: 391.6–1044.1 nm, resolution: 2.5 nm) and Camera 2 (Image-λ-N25E, wavelength range: 1044.1–2528.1 nm, resolution: 5.6 nm).

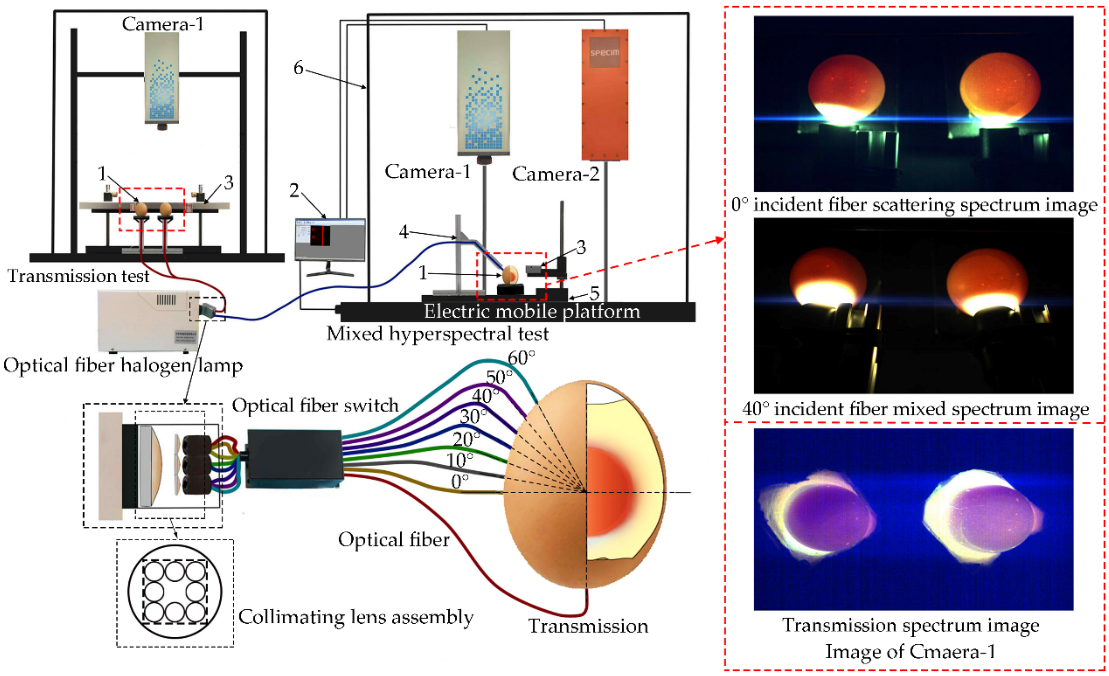
The reflection images of eggs were collected by the reflection hyperspectral imaging system (Figure 1). The light source of this system is a dome-uniform light source with a wavelength range of 50–2500 nm. The light source uniformly irradiate the egg on the electronically controlled moving platform. The reflected light of the egg is captured by the hyperspectral camera through the lens to obtain one-dimensional images and spectra. When the platform drives the egg to run continuously, continuous one-dimensional images and real-time spectra can be obtained. Note that the spectra are automatically recorded by computer software. Finally, we could obtain a three-dimensional data cube containing reflection image and spectral information.



1. Pink-shell egg; 2. Computer; 3. Black box; 4. Calibration whiteboard; 5. Sample table; 6. Dome-uniform light source; Camera 1. Visible near-infrared Camera; Camera 2. Short wave near-infrared Camera

Figure 1. Reflection hyperspectral imaging system

The scattering, transmission, and mixed hyperspectral images of eggs were collected by optical fiber hyperspectral imaging system (Figure 2). The light source of the system is an optical fiber halogen lamp (LG-150B, wavelength range: 400–2500 nm). The incident angle of the fiber can be adjusted to collect the corresponding types of hyperspectral images. The scattering hyperspectral images are collected as the incident angle is 0°. The mixed hyperspectral images are collected as the incident angles are 10°, 20°, 30°, 40°, 50° and 60°, respectively. The transmission hyperspectral images are collected as the light of fiber is shot directly under the egg. As the angle is selected, the platform drives the sample to move continuously to obtain continuous one-dimensional images and real-time spectral information. Finally, we could obtain a three-dimensional data cube including scattering, transmission, and mixed images and spectral information.



1. Pink-shell egg 2. Computer 3. Calibration whiteboard 4. Optical fiber fixed metal frame 5. Sample table 6. Black box; Camera 1. Visible near-infrared Camera; Camera 2. Short wave near-infrared Camera

Figure 2. Optical fiber hyperspectral imaging system

2.3. Data acquisition and correction

The equipment should be prepared before testing. The detection system should be warmed up for 30 minutes. The height of "Camera 1" is set to 10 cm, the exposure time is 7 ms by adjustment and comparison. The height of "Camera 2" is set to 25 cm, and the exposure time is 9 ms. The conveyor belt speed is 0.36 cm/sec. The hyperspectral images were collected by the following ways. 10 eggs were randomly selected from the data group every day, and the larger end of the eggs (with air chamber) was placed upward under the dome uniform light source to obtain the reflection hyperspectral images. Then they were placed in the transmission light and the optical fiber light sources with incident angles of 0°, 10°, 20°, 30°, 40°, 50° and 60° to obtain the scattering, transmission and mixed hyperspectral of the eggs. The tests are repeated and lasted for 28 days. The collected hyperspectral images are corrected in black and white because of the influence of dark current or uneven illumination on the experiment [16]. It was corrected by using SpecVIEW software established in the system and equation 1.

$$R = \frac{I_0 - I_b}{I_w - I_b} \quad (1)$$

Where, R is the corrected spectral image, I_0 is the original spectral image, I_w is the total reflection image of polyfluortetraethylene plate, I_b is the all-black image by coving the lens.

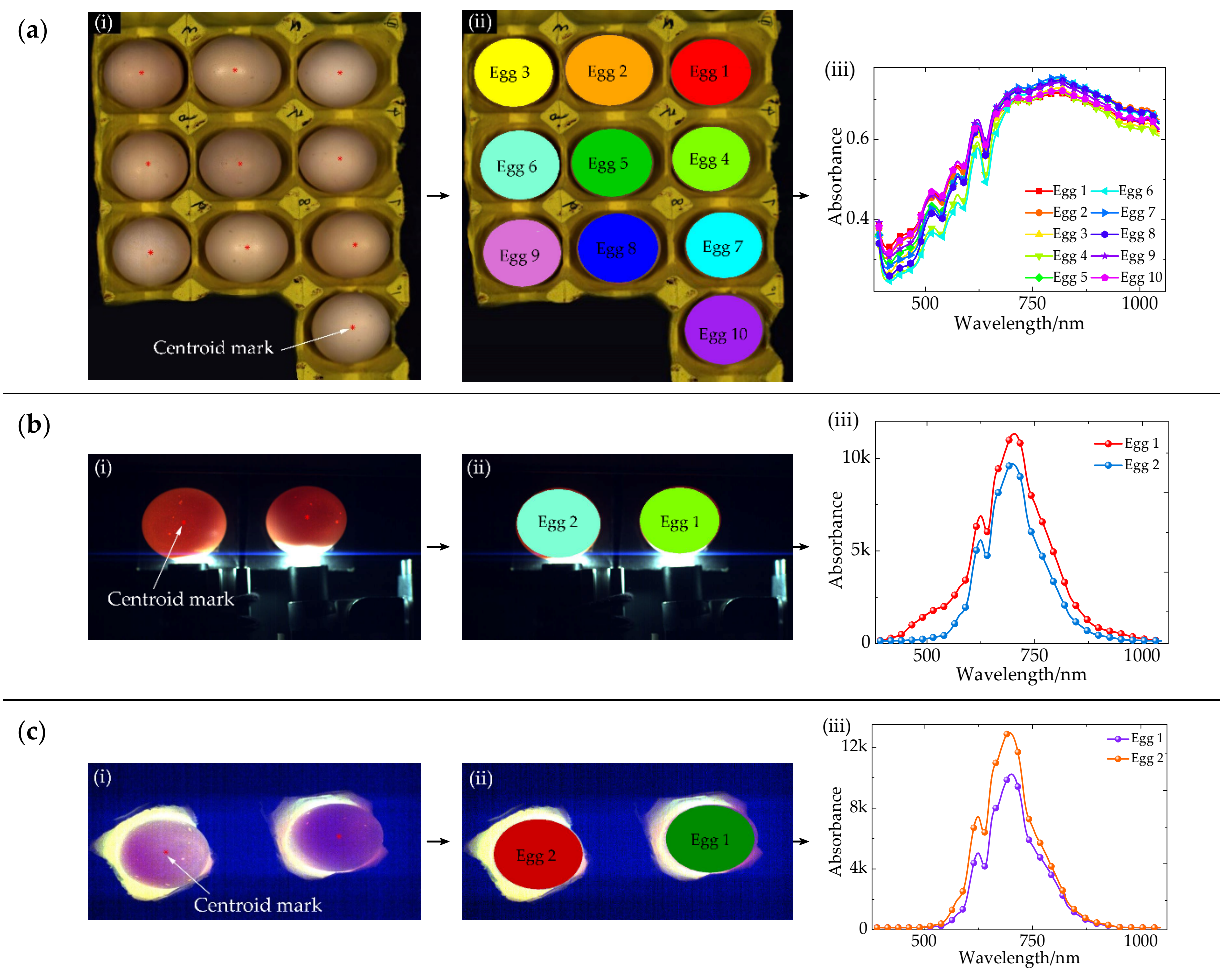
2.4. Automatic ROI extraction

Step 1 ROI mask

The images (R: 650, G: 550, B: 450) were exported by the software ENVI 4.8. The images were extracted by using MATLAB. They are binarized and then operated by threshold segmentation, expansion and erosion. Subsequently, their centroids were extracted and marked. According to the ellipse formula, we take the centroid as the center of the original image, and use the long axis and short axis parameters to fit and expand the ellipse image to extract the ROI (Regions of interest) mask.

Step 2 automatically extract the ROI of spectra

The positions of egg in the mask image are extracted by the cell-counting algorithm. The corresponding ROI of eggs are determined and numbered by the settings of their mask images. These images are import into ENVI. The average spectrum of a single ROI is used as the spectrum of an egg. The detailed processed are shown in Figure 3.



(i) Centroid mark; (ii) ROI; (iii) Original hyperspectral

Figure 3. ROI extraction process: (**a**) Reflection, (**b**) Transmission, (**c**) Mixed hyperspectral

2.5. Determination of Haugh unit

5 eggs were randomly selected from the calibration group every day, and were numbered and weighed. Their shells were broken gently, the height of protein were measured at 3 different points of 1 cm from the edge of their yolks. The three points were selected as far as possible, the average height were used as the protein height of an egg. The Haugh units of the 5 eggs are calculated by the equation 2, and their average value are used as the egg freshness of the day. [17]

$$HU=100\times lg(h+7.57-1.7*w^{0.37}) \tag{2}$$

Where, HU is Haugh unit of an egg, h (mm) is the average protein height of the three points; w (g) is the weight of an egg.

2.6. Spectrum processing method

It is necessary to preprocess the original spectra due to the uneven intensity of light sources at different wavelengths and the influence of instrument noise. In this paper, the spectra are processed by ten pre-processing methods, including Multiplicative Scatter Correction (MSC) [18], Standardized Normal Variate (SNV) [19], Normalization[20], Autoscales [21], Mean Centering (MC)[22], Moving-Average Method (MA) [23], Detrend Fluctuation Analysis (Detrend) [24], Savitsky-Golay Smoothing (SG) [25], Savitsky-Golay- First Derivative (SG-FD) [26] and Savitsky-Golay- Second Derivative (SG-SD) [27]. To reduce calculation and increase calculation speed, Competitive Adaptive Reweighted Sampling (CARS) [28], Principal Components Analysis (PCA) [29] and Successive Projections Algorithm (SPA) [30] are preferable to extract feature wavelengths to reduce the dimensionality. The preprocessed data set is used to extract feature wavelengths and used as the final sample. Then, 71.43% of the samples are randomly selected as the training set, and the remaining 28.57% 0% as the test set. We compared the prediction of egg freshness by the following six models, including Support Vector Machine (SVM) [31], K-Nearest Neighbor (KNN) [32], Random Forest (RF) [33], Naive Bayes (NB) [34], Discriminant Analysis Classifier (DAC) [35], Latent Dirichlet Allocation (LDA) [36]. In order to further improve the accuracy and the generalization ability of the egg freshness classification model, multiple weak classifiers are merged into a strong classifier by stacking ensemble learning [37].

3. Guided Filtering

3.1. Determination of egg Haugh unit

5 eggs were selected randomly every day to measure their Haugh units, and the units of 140 eggs were measured within 28 days. The units decrease linearly with time (Figure 4), and they fits well with the equation 3. Their detail distribution are shown in Table 1.

$$y = 85.70\text{-}1.75\ x \tag{3}$$

It shows that the Haugh units range from 33.4 to 84.5, thus these eggs are edible. Their units are 84.5–72, 70.5–61.5, 59.8–49.0, and 47.2–33.4 in the first, second, third, and fourth week, and their freshness are classified as Grade AA, A, B_1, and B_2, respectively. After the fourth week, their units are below 30 and classified as Grade C due to their Haugh units gradually decrease. These eggs are easy to distinguish due to their obvious spoilage and unpleasant smell deterioration, so they will not be discussed in this article.

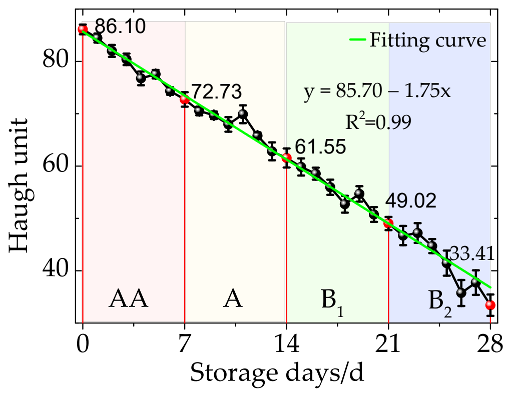


Figure 4. Haugh unit of eggs versus time

Table 1. Distribution of egg Hastelloy

Freshness	weeks	Max	Min	Average	Standard deviation
AA	1	84.463	72.731	78.316	4.198
A	2	70.527	61.546	66.893	3.602

B_1	3	59.824	49.019	54.493	3.962
B_2	4	47.202	33.408	40.991	5.492

3.2 Spectral preprocessing

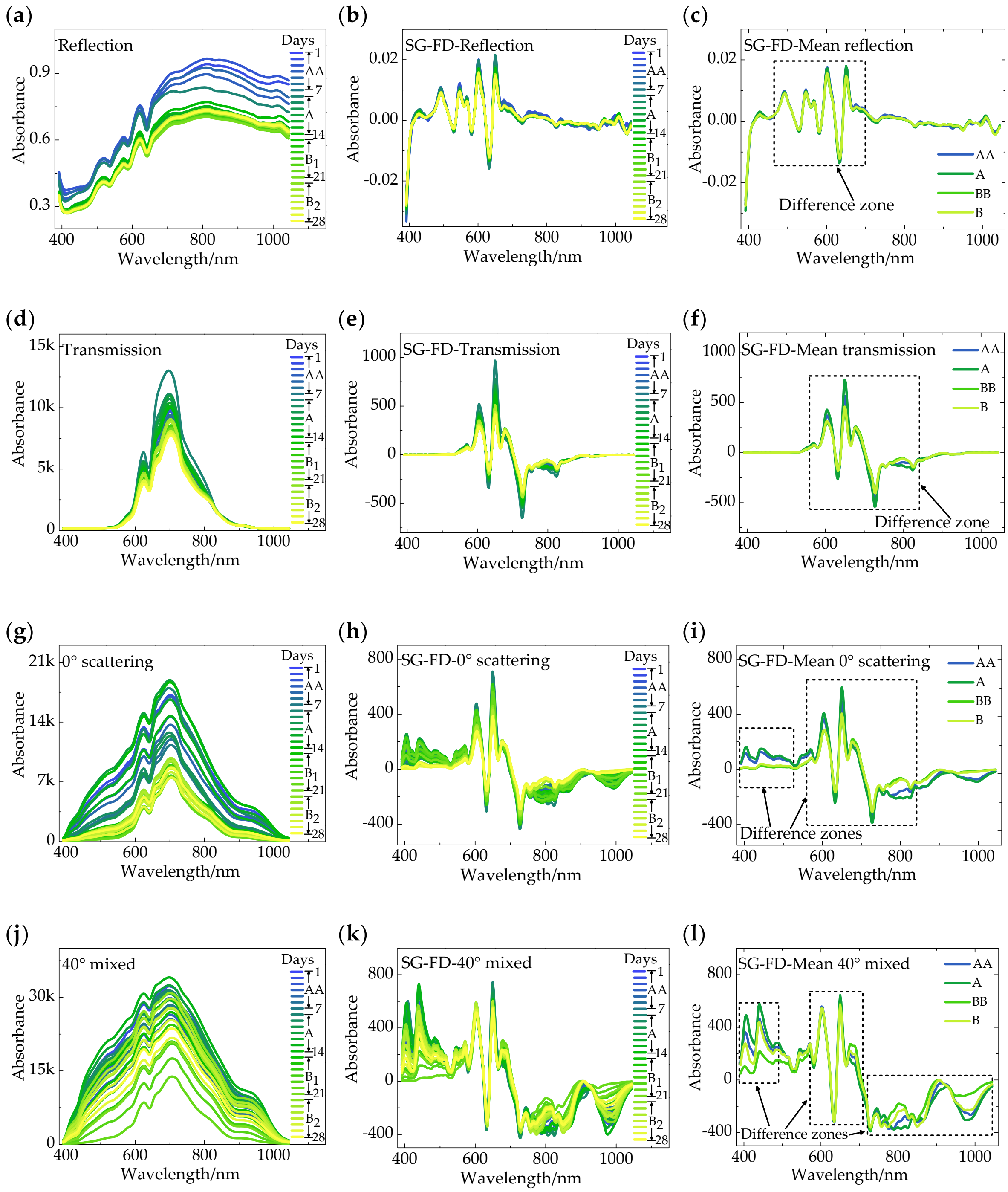


Figure 5. SG-FD pretreatment: (a), (d), (g) and (j) the average hyperspectral of eggs per day from 1 to 28 days as incident light are reflection, transmission, 0° and 40° scattering, respectively. (b), (e), (h) and (j) corresponding spectra after SG-FD treatment. (c), (f), (g) and (k) the average spectra of Grade AA, A, B_1 and B_2 after SG-FD treatment, respectively.

After the previous preprocessing, the freshness is modeled by feature extraction. This paper only discusses the modeling results of the hyperspectral data of the fiber light source with four typical incident angles, uniform reflection, transmission, 0° and 40°, due to the large amount of data.

 doi:10.20944/preprints202008.0378.v1

The original spectra contain a lot of information about the freshness of eggs, however, it is impossible to find the law directly (Figure 5a, 5d, 5h and 5g). The spectra have obvious noise, which will interfere with the later extraction of feature wavelength and modeling, and reduce the accuracy of the prediction model. Therefore, the original spectra should be preprocessed separately. SG is an algorithm of polynomial smoothing and weighted average of moving windows based on the principle of least squares. While, the main idea of FD is to obtain the first derivative of the spectrum, thereby amplifying the difference between different spectrum. Herein, the original spectra are treated by SG-FD (Figure 5b, 5e, 5h and 5k). We obtain the average of the four Grade, AA, A, B_1 and B_2, after the SG-FD treatment (Figure 5c, 5f, 5i and 5l). The spectral differences of eggs with different freshness are mainly distributed in the wavelength bands of 400–600, 550–800, 550–800 and 400–1000 nm in uniform reflection, transmission, 0° scattering and 40° mixed spectra, respectively.

3.3 Feature wavelength extraction and model establishment

In our experiment, we use PCA, CARS and SPA to extract the feature wavelengths and reduce the redundancy of the full-band original spectra. It can eliminate irrelevant information, optimize effective information, and establish low-dimensional data models. Finally, different classification models are established according to the feature wavelength, and the best model is obtained by comparative analysis.

3.3.1 Model based on PCA

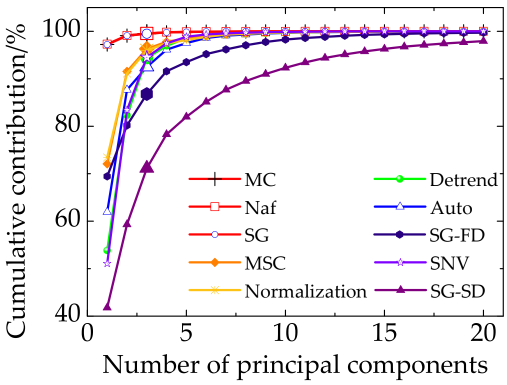


Figure 6. Cumulative contribution rate of the top 20 principal components in 0° incident light

Table 2. Cumulative contribution rate of the first three principal components

Pretreatment method	Cumulative contribution rate /% (the first three principal components)			
	Reflection	Transmission	0°	40°
MSC	89.16	94.97	96.31	95.64
SNV	89.01	96.84	92.50	88.63
Normalization	**95.58**	**95.11**	**95.98**	**95.05**
Auto	89.01	96.84	92.50	88.63
MC	**99.31**	**90.70**	**99.51**	**99.15**
MA	**99.30**	**91.49**	**99.54**	**99.25**
Detrend	**91.63**	**98.99**	**94.38**	**90.89**
SG	**99.33**	**90.87**	**99.48**	**99.06**
SG-FD	79.72	76.44	86.79	80.21
SG-SD	**90.26**	**98.15**	**91.14**	**96.73**

PCA analysis based on preprocessed data. We take the various pre-processing methods of scattering hyperspectral at the 0° incident light as an example, calculate the cumulative contribution of the first 20 principal components (Figure 6). The first 3 principal components have the highest contribution They were selected as feature component. Meanwhile, the cumulative contribution of different pretreatments are shown in Table 2. It can be seen that the cumulative contribution of the

 doi:10.20944/preprints202008.0378.v1

first 3 components for Normalization, MC, MA, Detrend, SG, and SG-SD are above 90%. Therefore, the first 3 components of these pretreatments are selected as the new coordinate system to reduce dimension of the original spectra and extract the feature wavelengths. Then, we establish LIBSVM, DCA, LDA, KNN, RF and NB models to calculate the accuracy of training set and test set, respectively (Table 3). The results show that the overall accuracy of PCA-based feature wavelength extraction is not high. Among which LDA modeling has the highest classification accuracy of only 82.5%.

Table 3. PCA-based modeling results

M.	P.	Training set prediction accuracy /%				Prediction set training accuracy /%			
		R.	T.	0 °	40 °	R.	T.	0 °	40 °
LIBSVM	Norm.	69.50	81.50	78.00	79.00	53.75	66.25	62.50	66.25
	MC	61.00	88.00	51.50	51.50	43.75	72.50	46.25	43.75
	MA	60.00	89.00	51.50	51.00	42.50	72.50	45.00	43.75
	Detrend	71.50	64.50	81.50	87.50	56.25	52.50	57.50	61.25
	SG	60.00	89.00	51.50	51.00	43.75	76.25	46.25	43.75
	SG-SD	98.00	97.50	96.00	97.50	**80.00**	78.75	75.00	76.25
DAC	Norm.	65.50	82.50	79.00	80.50	53.75	65.00	63.75	63.75
	MC	57.50	91.50	54.50	49.00	42.50	66.25	46.25	43.75
	MA	59.00	91.00	54.50	49.00	42.50	66.25	46.25	43.75
	Detrend	77.50	66.00	79.00	86.00	60.00	56.25	61.25	58.75
	SG	57.50	91.00	54.50	49.00	42.50	66.25	46.25	43.75
	SG-SD	97.00	96.50	93.50	94.00	77.50	**80.00**	71.25	72.50
LDA	Norm.	61.00	77.50	74.00	73.00	52.50	58.75	58.75	47.50
	MC	43.00	89.50	48.50	39.00	41.25	66.25	45.00	38.75
	MA	43.50	90.00	48.00	39.50	40.00	67.50	45.00	37.50
	Detrend	72.00	67.00	71.00	83.50	60.00	57.50	62.50	57.50
	SG	43.50	89.50	48.00	39.00	41.25	68.75	45.00	38.75
	SG-SD	97.00	91.50	95.00	96.00	**80.00**	**82.50**	77.50	77.50
KNN	Norm.	100.00	100.00	100.00	100.00	60.00	73.75	63.75	56.25
	MC	100.00	100.00	100.00	100.00	58.75	65.00	47.50	48.75
	MA	100.00	100.00	100.00	100.00	57.50	65.00	47.50	48.75
	Detrend	100.00	100.00	100.00	100.00	77.50	63.75	66.25	57.50
	SG	100.00	100.00	100.00	100.00	57.50	65.00	47.50	48.75
	SG-SD	100.00	100.00	100.00	100.00	78.75	72.50	73.75	**80.00**
RF	Norm.	100.00	100.00	100.00	100.00	63.75	71.25	58.75	61.25
	MC	100.00	100.00	100.00	100.00	53.75	75.00	46.25	45.00
	MA	100.00	100.00	100.00	100.00	53.75	68.75	46.25	43.75
	Detrend	100.00	100.00	100.00	100.00	73.75	53.75	61.25	60.00
	SG	100.00	100.00	100.00	100.00	55.00	73.75	45.00	43.75
	SG-SD	100.00	100.00	100.00	100.00	75.00	71.25	67.50	**80.00**
NB	Norm.	62.50	72.00	69.50	78.00	50.00	62.50	47.50	57.50
	MC	62.50	73.50	55.00	49.50	51.25	63.75	48.75	42.50
	MA	62.50	73.00	55.00	49.50	51.25	61.25	50.00	42.50
	Detrend	78.00	63.00	71.00	81.00	63.75	57.50	51.25	56.25
	SG	62.50	75.00	55.00	49.50	51.25	61.25	48.75	42.50
	SG-SD	93.00	88.00	90.00	91.00	70.00	66.25	70.00	76.25

[3] **M.**: Model; **P.**: Pretreatment; **R.**: Reflection; **T.**: Transmission; **0°**: 0° incident light; **40°**: 40° incident light.

3.3.2 Model based on SPA

Successive projections algorithm (SPA) can eliminate collinear redundancy to find the wavelength segment with the minimum collinear information and represent the maximum information of the sample. In this experiment, the number of wavelengths selected by SPA is set to range from 5 to 30, and the step length is 1. Then, we iterate the data and select the wavelength with the largest projection phasor as the feature wavelength combination. Meanwhile, the RMSE of different combinations is calculated by linear regression until the feature wavelength combination corresponding to the minimum RMSE is obtained. The SPA feature wavelength is extracted from the preprocessed data of SG-FD as the incident angle of 0°. The result shows that the best RMSE = 0.58 as the feature wavelength is 22.

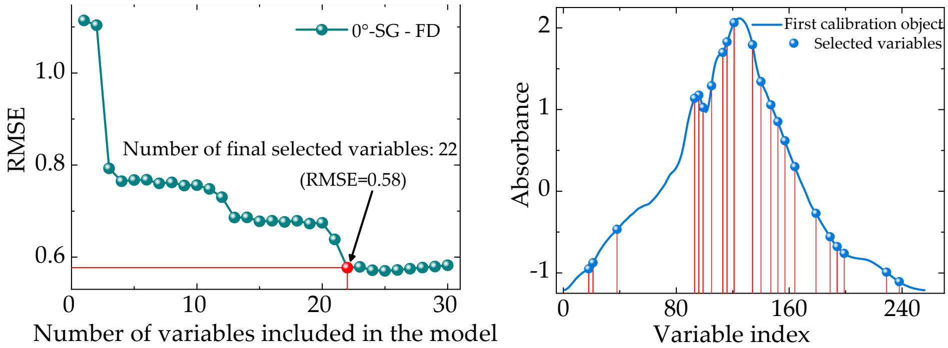


Figure 7. Number of variables in the 0° incident light model

The number of feature wavelength is different extracted by different preprocessing methods (Table 4). Subsequently, LIBSVM, DCA, LDA, KNN, RF and NB models were established to obtain the accuracy of the training set and the test set (Table 5). By comparing Table 3 and 5, it can be concluded that the overall accuracy of feature wavelength extraction based on SPA is higher than that of PCA. Meanwhile, the pretreatment by MSC, SNV, Auto, MC with the classification of DAC model have higher accuracy. 0° incidence angle by MSC-SPA-DAC has the highest accuracy (96.25%), while that of reflection incidence by SG-SPA-SG is low (81.25%). This is consistent with that the light of scattering has more internal information of the egg than that of reflection.

Table 4. Number of feature wavelengths extracted after SPA processing

P.	MSC	SNV	Norm.	Auto	MC	MA	Detrend	SG	SG-FD	SG-SD
R.	17	10	16	10	11	19	12	18	16	12
T.	20	18	17	18	12	15	21	20	12	6
0°	22	22	16	22	11	18	19	19	22	15
40°	35	48	47	26	24	16	26	35	35	44

[4] **N**: Number of feature wavelengths; **P.**: Pretreatment; **R.**: Reflection; **T.**: Transmission.

Table 5. SPA-based modeling results

M.	P.	Training set prediction accuracy /%				Prediction set training accuracy /%			
		R.	**T.**	**0 °**	**40 °**	**R.**	**T.**	**0 °**	**40 °**
	MSC	84.00	89.50	88.00	87.50	60.00	68.75	70.00	70.00
	SNV	67.00	92.00	91.50	97.00	47.50	78.75	73.75	71.25
	Norm.	75.50	92.00	81.50	87.00	57.50	68.75	65.00	71.25
	Auto	67.00	92.00	91.50	97.00	47.50	78.75	73.75	71.25
	MC	59.00	76.50	58.00	64.00	43.75	62.50	42.50	48.75
LibSVM	MA	61.00	82.50	63.00	61.50	46.25	62.50	41.25	48.75
	Detrend	80.50	87.00	88.50	92.00	60.00	67.50	73.75	61.25
	SG	58.00	86.50	65.00	65.00	46.25	70.00	47.50	52.50
	SG-FD	98.00	94.50	95.50	96.50	80.00	75.00	78.75	78.75
	SG-SD	94.00	84.50	96.50	95.50	76.25	58.75	77.50	75.00
	MSC	94.50	100.00	100.00	99.00	71.25	**91.25**	**96.25**	85.00

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2020 doi:10.20944/preprints202008.0378.v1

	SNV	90.00	100.00	100.00	100.00	68.75	**90.00**	**93.75**	**92.50**
	Norm.	87.00	99.50	99.00	99.50	62.50	**90.00**	**91.25**	**91.25**
	Auto	90.00	100.00	100.00	100.00	68.75	**90.00**	**93.75**	**92.50**
DAC	MC	97.50	96.50	98.00	99.00	80.00	75.00	83.75	87.50
	MA	97.50	99.00	99.50	100.00	77.50	81.25	86.25	**91.25**
	Detrend	95.00	98.50	99.00	99.00	77.50	86.25	85.00	**91.25**
	SG	98.00	97.50	99.00	99.00	77.50	81.25	85.00	**90.00**
	SG-FD	97.00	97.50	98.00	98.50	77.50	77.50	**92.50**	**90.00**
	SG-SD	91.50	87.00	95.50	96.50	66.25	63.75	85.00	82.50
	MSC	90.00	82.00	74.00	90.00	72.50	76.25	70.00	71.25
	SNV	75.50	81.00	89.00	86.50	61.25	71.25	83.75	67.50
	Norm.	94.50	74.50	78.50	95.00	75.00	72.50	85.00	77.50
	Auto	75.50	81.00	89.00	86.50	61.25	71.25	83.75	67.50
LDA	MC	89.50	62.50	86.00	92.50	71.25	50.00	75.00	78.75
	MA	97.50	60.00	86.00	91.50	77.50	61.25	77.50	76.25
	Detrend	89.50	80.50	85.00	88.50	73.75	66.25	76.25	80.00
	SG	97.00	75.00	71.50	94.00	**81.25**	60.00	66.25	78.75
	SG-FD	93.50	66.50	81.50	90.50	68.75	50.00	77.50	75.00
	SG-SD	86.50	45.50	96.50	91.00	66.25	32.50	78.75	65.00
	MSC	100.00	100.00	100.00	100.00	66.25	77.50	63.75	62.50
	SNV	100.00	100.00	100.00	100.00	65.00	77.50	65.00	71.25
	Norm.	100.00	100.00	100.00	100.00	63.75	73.75	62.50	63.75
	Auto	100.00	100.00	100.00	100.00	65.00	77.50	65.00	71.25
KNN	MC	100.00	100.00	100.00	100.00	63.75	68.75	57.50	55.00
	MA	100.00	100.00	100.00	100.00	57.50	63.75	52.50	53.75
	Detrend	84.00	89.50	88.00	87.50	60.00	68.75	70.00	70.00
	SG	67.00	92.00	91.50	97.00	47.50	78.75	73.75	71.25
	SG-FD	75.50	92.00	81.50	87.00	57.50	68.75	65.00	71.25
	SG-SD	67.00	92.00	91.50	97.00	47.50	78.75	73.75	71.25
	MSC	59.00	76.50	58.00	64.00	43.75	62.50	42.50	48.75
	SNV	61.00	82.50	63.00	61.50	46.25	62.50	41.25	48.75
	Norm.	80.50	87.00	88.50	92.00	60.00	67.50	73.75	61.25
	Auto	58.00	86.50	65.00	65.00	46.25	70.00	47.50	52.50
RF	MC	98.00	94.50	95.50	96.50	80.00	75.00	78.75	78.75
	MA	94.00	84.50	96.50	95.50	76.25	58.75	77.50	75.00
	Detrend	94.50	100.00	100.00	99.00	71.25	**91.25**	**96.25**	85.00
	SG	90.00	100.00	100.00	100.00	68.75	**90.00**	**93.75**	**92.50**
	SG-FD	87.00	99.50	99.00	99.50	62.50	**90.00**	**91.25**	**91.25**
	SG-SD	90.00	100.00	100.00	100.00	68.75	**90.00**	**93.75**	**92.50**
	MSC	97.50	96.50	98.00	99.00	80.00	75.00	83.75	87.50
	SNV	97.50	99.00	99.50	100.00	77.50	81.25	86.25	**91.25**
	Norm.	95.00	98.50	99.00	99.00	77.50	86.25	85.00	**91.25**
	Auto	98.00	97.50	99.00	99.00	77.50	81.25	85.00	**90.00**
NB	MC	97.00	97.50	98.00	98.50	77.50	77.50	**92.50**	**90.00**
	MA	91.50	87.00	95.50	96.50	66.25	63.75	85.00	82.50
	Detrend	90.00	82.00	74.00	90.00	72.50	76.25	70.00	71.25
	SG	75.50	81.00	89.00	86.50	61.25	71.25	83.75	67.50
	SG-FD	94.50	74.50	78.50	95.00	75.00	72.50	85.00	77.50
	SG-SD	75.50	81.00	89.00	86.50	61.25	71.25	83.75	67.50

[5] **M.**: Model; **P.**: Pretreatment; **R.**: Reflection; **T.**: Transmission; 0°: 0° incident light; 40°: 40° incident light.

3.3.3 Model based on CARS

Competitive Adaptive Reweighted Sampling (CARS) is based on the principle of "survival of the fittest" in Darwin's theory of evolution. In order to reduce the dimensionality, partial least squares are used to select the spectral value with a larger regression coefficient, and the value with a smaller one is eliminated to select some feature wavelengths for representing the full spectral information. After this preprocessing, the dimensionality of the data is effectively reduced. In this paper, we reduced the dimensionality of the preprocessed spectrum by CARS and sample the eggs by Monte Carlo. The sampling time of Monte Carlo was set to 100, PLS model was established by using 5-fold cross-validation. Subsequently, the 0° incident light was taken as an example to extract the process of the feature wavelengths after SG-FD preprocessing (Figure 8).

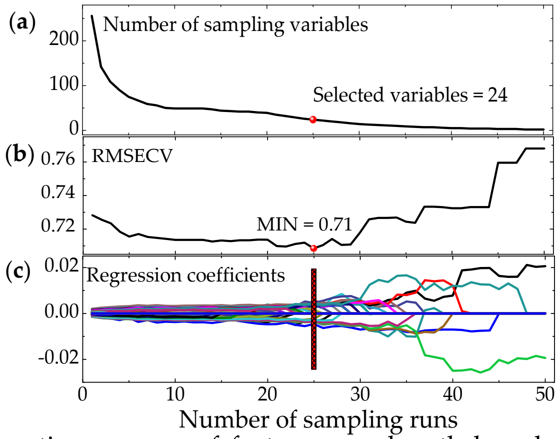


Figure 8. The extraction process of feature wavelength based on CARS at 0° incident source: (**a**) Number of sampling variables; (**b**) RMSECV; (**c**) Regression coefficient path

The number of retained wavelengths decreases slowly after starting to decrease rapidly as the sampling frequency increases. RMSECV decreases slowly as the number of sampling runs ranging from 0 to 24, indicating that the eliminated wavelength has little influence on RMSECV. However, it increases significantly as the number exceeds 24, indicating that the feature wavelengths have been deleted. Therefore, the number of extracted feature wavelengths is 24. Similarly, the number preprocessed by other methods can be extracted (Table 6).

Table 6. Number of feature wavelengths extracted after CARS processing

P.	MSC	SNV	Norm.	Auto	MC	MA	Detrend	SG	SG-FD	SG-SD
R	16	43	26	32	32	14	46	32	40	46
T	24	35	29	24	20	28	23	27	24	36
0°	28	44	26	16	16	16	20	22	24	16
40°	16	24	16	24	16	21	19	20	19	15

[6] **N**: Number of feature wavelengths; **P.**: Pretreatment; **R.**: Reflection; **T.**: Transmission.

Subsequently, the egg freshness classification models are established by LIBSVM, DCA, LDA, KNN, RF and NB. The results show that the prediction accuracy of egg freshness based on CARS is better than that of SPA and PCA, and its accuracy rate reaches 90%. Among of which the DAC and KNN models have the highest accuracy, both of them reaches 95%.

Table 7. Accuracy of CARS model

M.	P.	Training set prediction accuracy /%				Prediction set training accuracy /%			
		R.	**T.**	**0 °**	**40 °**	**R.**	**T.**	**0 °**	**40 °**
LibSVM	MSC	83.50	93.00	92.50	94.00	58.75	75.00	70.00	73.75
	SNV	93.00	94.00	94.50	97.00	68.75	80.00	72.50	73.75
	Norm.	70.00	92.50	90.50	93.50	53.75	78.75	71.25	72.50
	Auto	86.50	92.00	93.00	93.50	62.50	78.75	70.00	71.25
	MC	59.50	84.00	61.00	65.50	48.75	70.00	43.75	50.00
	MA	62.00	89.00	65.00	62.50	47.50	71.25	52.50	46.25
	Detrend	88.50	88.00	92.00	91.50	66.25	67.50	76.25	66.25
	SG	57.50	88.00	66.00	70.50	46.25	76.25	51.25	53.75
	SG-FD	98.00	97.00	97.00	96.50	80.00	78.75	83.75	76.25
	SG-SD	98.00	97.00	98.50	100.00	80.00	81.25	**90.00**	81.25

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2020 doi:10.20944/preprints202008.0378.v1

DAC	MSC	94.00	99.50	100.00	100.00	73.75	87.50	**91.25**	**95.00**
	SNV	98.00	100.00	100.00	100.00	81.25	**90.00**	**95.00**	**90.00**
	Norm.	97.50	100.00	100.00	100.00	80.00	**92.50**	**93.75**	88.75
	Auto	98.00	100.00	100.00	100.00	76.25	**91.25**	**95.00**	88.75
	MC	98.00	99.00	99.50	100.00	80.00	80.00	86.25	**90.00**
	MA	98.00	99.00	100.00	99.00	78.75	78.75	**93.75**	87.50
	Detrend	98.00	99.50	99.50	99.50	80.00	88.75	**93.75**	**93.75**
	SG	98.00	98.00	99.50	100.00	78.75	85.00	**95.00**	**92.50**
	SG-FD	98.00	99.00	100.00	99.50	82.50	78.75	**92.50**	**92.50**
	SG-SD	98.00	99.00	100.00	100.00	78.75	82.50	**92.50**	**92.50**
LDA	MSC	88.00	82.00	87.00	98.50	70.00	65.00	71.25	62.50
	SNV	97.00	99.00	99.00	97.50	81.25	72.50	76.25	71.25
	Norm.	87.50	89.50	90.00	100.00	80.00	73.75	62.50	65.00
	Auto	89.50	69.50	85.50	82.00	78.75	55.00	65.00	62.50
	MC	95.00	63.50	89.00	90.50	83.75	62.50	77.50	73.75
	MA	83.00	61.50	83.50	73.00	65.00	56.25	67.50	56.25
	Detrend	100.00	78.50	85.00	92.50	86.25	66.25	58.75	78.75
	SG	95.00	59.50	86.50	88.00	82.50	38.75	65.00	67.50
	SG-FD	98.50	78.00	83.50	96.00	83.75	57.50	61.25	73.75
	SG-SD	98.00	79.00	90.50	97.00	78.75	65.00	76.25	71.25
KNN	MSC	100.00	100.00	100.00	100.00	73.75	87.50	**91.25**	**95.00**
	SNV	100.00	100.00	100.00	100.00	81.25	**90.00**	**95.00**	**90.00**
	Norm.	100.00	100.00	100.00	100.00	80.00	**92.50**	**93.75**	88.75
	Auto	100.00	100.00	100.00	100.00	76.25	**91.25**	**95.00**	88.75
	MC	100.00	100.00	100.00	100.00	80.00	80.00	86.25	**90.00**
	MA	100.00	100.00	100.00	100.00	78.75	78.75	**93.75**	87.50
	Detrend	83.50	93.00	92.50	94.00	58.75	75.00	70.00	73.75
	SG	93.00	94.00	94.50	97.00	68.75	80.00	72.50	73.75
	SG-FD	70.00	92.50	90.50	93.50	53.75	78.75	71.25	72.50
	SG-SD	86.50	92.00	93.00	93.50	62.50	78.75	70.00	71.25
RF	MSC	59.50	84.00	61.00	65.50	48.75	70.00	43.75	50.00
	SNV	62.00	89.00	65.00	62.50	47.50	71.25	52.50	46.25
	Norm.	88.50	88.00	92.00	91.50	66.25	67.50	76.25	66.25
	Auto	57.50	88.00	66.00	70.50	46.25	76.25	51.25	53.75
	MC	98.00	97.00	97.00	96.50	80.00	78.75	83.75	76.25
	MA	98.00	97.00	98.50	100.00	80.00	81.25	**90.00**	81.25
	Detrend	94.00	99.50	100.00	100.00	73.75	87.50	**91.25**	**95.00**
	SG	98.00	100.00	100.00	100.00	81.25	**90.00**	**95.00**	**90.00**
	SG-FD	97.50	100.00	100.00	100.00	80.00	**92.50**	**93.75**	88.75
	SG-SD	98.00	100.00	100.00	100.00	76.25	**91.25**	**95.00**	88.75
NB	MSC	98.00	99.00	99.50	100.00	80.00	80.00	86.25	**90.00**
	SNV	98.00	99.00	100.00	99.00	78.75	78.75	**93.75**	87.50
	Norm.	98.00	99.50	99.50	99.50	80.00	88.75	**93.75**	**93.75**
	Auto	98.00	98.00	99.50	100.00	78.75	85.00	**95.00**	**92.50**
	MC	98.00	99.00	100.00	99.50	82.50	78.75	**92.50**	**92.50**
	MA	98.00	99.00	100.00	100.00	78.75	82.50	**92.50**	**92.50**
	Detrend	88.00	82.00	87.00	98.50	70.00	65.00	71.25	62.50
	SG	97.00	99.00	99.00	97.50	81.25	72.50	76.25	71.25
	SG-FD	87.50	89.50	90.00	100.00	80.00	73.75	62.50	65.00
	SG-SD	89.50	69.50	85.50	82.00	78.75	55.00	65.00	62.50

[7] **M.**: Model; **P.**: Pretreatment; **R.**: Reflection; **T.**: Transmission; **0°**: 0° incident light; **40°**: 40° incident light.

3.4 Best prediction model of egg freshness

The method presented in 3.3 is used for the nine different incident light modes to select their highest accuracy of egg freshness, respectively (Table 8). It shows that the overall model accuracy is extracted the feature wavelength of CARS, which is higher than PCA and SPA. Among of them, the weak classifiers DAC, KNN and PCA have the three highest accuracies. Moreover, the accuracy of

 doi:10.20944/preprints202008.0378.v1

MSC-SPA-DAC model (96.25%) is the highest as the incident light angle is 0°. The accuracy of MA-CARS-KNN at the 30°-incident light and MSC-CARS-DAC at the 40°-incident light are 95% and 95%, respectively. The model at mean reflection light and 60°-incident light have low accuracy, 86.25% and 87.5%, respectively. It indicate that the accuracy of the scattering hyperspectral model is higher than the other three. In addition, as the angle of incidence increases, the overall accuracy decreases.

Table 8. The highest accuracy of a model under different incident light

Incident light		**The best model**	**Accuracy / %**
Mean reflection light		Detrend-CARS-LDA	86.25
Optical fiber	transmission	Nomalization-CARS-DAC	92.50
	0 °	MSC-SPA-DAC	**96.25**
	10 °	MA-CARS-PCA	93.75
	20 °	SNV/Auto-SPA-DAC	92.50
	30 °	MA-CARS-KNN	95.00
	40 °	MSC-CARS-DAC	95.00
	50 °	Detrend/SG-CARS-DAC Detrend-SPA-DAC	91.25
	60 °	SG-FD-SPA-KNN	87.50

3.5 Egg freshness classification based on stacking ensemble learning

To further improve the accuracy of the model, several weak classifiers are combined into a strong classifier, and stacking ensemble learning (SEL) [27] is performed to improve the generalization ability of the classification model. A two-layer training structure of SEL is used to improve the accuracy and speed of model. The overall flow chart of stacking ensemble learning is shown in Figure 9. The first layer uses different classifiers to establish different meta-classifiers and integrates the prediction results of all meta-classifiers. Then, the integrated data set of the classifiers with high accuracy in the first layer is used as the input of the second layer. Finally, the second layer is trained with the best classifier.

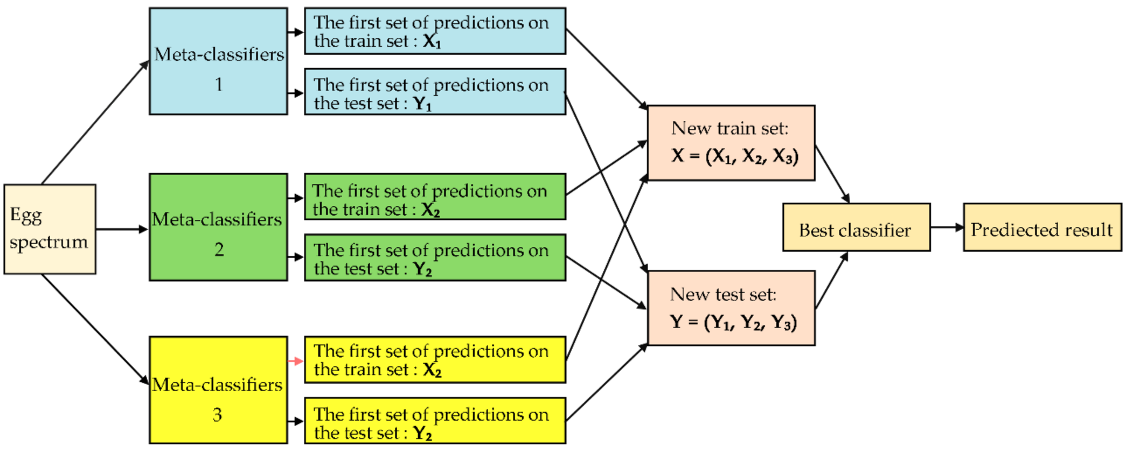


Figure 9. General flowchart for stacking ensemble learning

Therefore, in this experiment, three classifiers with the best model accuracy are selected to establish three meta-classifiers as the input of the second layer. The training and test set are predicted based on the idea of 5-fold cross validation in each meta-classifiers in order to prevent data leakage (Figure 10）. Finally, the new training and test set are used to establish the egg freshness classification model based on SEL.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 August 2020 doi:10.20944/preprints202008.0378.v1

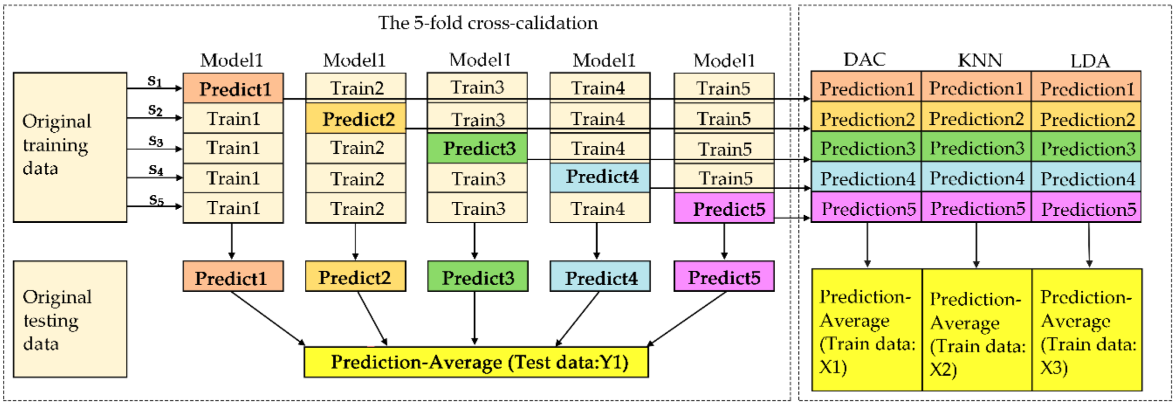


Figure 10. Training and prediction models for meta-models

The three classifiers, DAC, KNN and LDA, with the best accuracy are selected as the first layer. Meanwhile, the DAC model with the highest accuracy is selected as the second layer. Table 9 shows the results of the uniform reflection light source and transmission, 0° and 40° incident light sources.

Table 9. Modeling results of stacking ensemble learning

M.	P.	Training set prediction accuracy /%				Prediction set training accuracy /%			
		R.	T.	0°	40°	R.	T.	0°	40°
PCA	Norm.	69.50	82.50	79.00	80.50	68.75	78.75	70.00	68.75
	MC	62.50	91.00	55.00	51.00	61.25	73.75	53.75	53.75
	MA	78.00	67.00	81.50	87.50	78.75	72.50	72.50	73.75
	Detrend	62.50	91.00	55.00	51.00	78.75	86.25	75.00	77.50
	SG	97.50	91.00	89.50	93.50	83.75	75.00	75.00	81.25
	SG-SD	100.00	97.50	100.00	97.50	81.25	86.25	82.50	82.50
SPA	MSC	94.00	99.50	100.00	100.00	75.00	**93.75**	**100.00**	**90.00**
	SNV	98.00	100.00	100.00	100.00	71.25	**91.25**	**97.50**	**95.00**
	Norm.	97.50	100.00	100.00	100.00	77.50	**93.75**	**97.50**	**93.75**
	Auto	98.00	100.00	100.00	100.00	72.50	**93.75**	**95.00**	**95.00**
	MC	98.00	99.00	99.50	100.00	81.25	76.25	**90.00**	**92.50**
	MA	98.00	99.00	100.00	99.00	78.75	83.75	88.75	**95.00**
	Detrend	100.00	99.50	99.50	99.50	82.50	**92.50**	88.75	**96.25**
	SG	98.00	98.00	99.50	100.00	85.00	83.75	**91.25**	**92.50**
	SG-FD	98.50	99.00	100.00	99.50	82.50	78.75	**98.75**	**92.50**
	SG-SD	98.00	99.00	100.00	100.00	85.00	66.25	**90.00**	87.50
CARS	MSC	94.50	100.00	100.00	99.00	80.00	**90.00**	**95.00**	**96.25**
	SNV	90.00	100.00	100.00	100.00	87.50	**92.50**	**97.50**	**92.50**
	Norm.	94.50	99.50	99.00	99.50	85.00	**95.00**	**95.00**	**92.50**
	Auto	90.00	100.00	100.00	100.00	85.00	**95.00**	**98.75**	**93.75**
	MC	97.50	96.50	98.00	99.00	87.50	86.25	**90.00**	**93.75**
	MA	97.50	99.00	99.50	100.00	82.50	81.25	**95.00**	**90.00**
	Detrend	95.00	98.50	99.00	99.00	**88.75**	**90.00**	**97.50**	**95.00**
	SG	98.00	97.50	99.00	99.00	86.25	87.50	**98.75**	**96.25**
	SG-FD	98.00	97.50	98.00	98.50	86.25	83.75	**98.75**	**95.00**
	SG-SD	94.00	87.00	96.50	96.50	81.25	86.25	**97.50**	**93.75**

[7] **M.**: Model; **P.**: Pretreatment; **R.**: Reflection; **T.**: Transmission; **0°**: 0° incident light; **40°**: 40° incident light.

SEL modeling can improve the accuracy of the egg freshness models under different incident light. The 0°incident light source based on MSC-SPA can be increased from 96.25% to 100% (Table 10).

Table 10. The highest accuracy of the best model under different incident modes

Incident light		The best model	Accuracy / %
Mean reflection light		Detrend-CARS	88.75
	transmission	Normalization/Auto-CARS	95.00
	0 °	MSC-SPA	**100.00**
	10 °	SG-CARS	96.25
Optical fiber	20 °	SNV-SPA、MSC-CARS	95.00
	30 °	MA-CARS	96.25
	40 °	SG/MSC -CARS、Detrend-SPA	96.25
	50 °	SG-CARS、MA/Detrend-CARS	93.75
	60 °	SG-FD-SPA	90.00

The highest accuracy of the best model is different under different incident angles (Figure 11). The accuracy at the 0° incident light (100%) is the highest. Their accuracy are almost linearly reduced from 100% to 90% as the incident angle increases from 0° to 60°. The accuracy of the transmission and reflection incident model are 92.5% and 87.5%， respectively. These indicate that the incident angle has an important influence on the accuracy of a model.

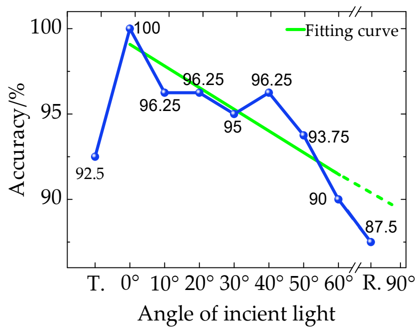


Figure 11. Classification accuracy-incident light line chart

4. Discussion

The accuracy of the non-destructive detection model for egg freshness based on hyperspectral can be improved by stacking ensemble learning. The learning is to use the output results of a series of models (base-model) as the input features of other models. This method realizes the stacking of models, that is, the outputs of the first layer model are used as the inputs of the second layer model. In operation, we need to pay attention to no leakage when combining the output of the first layer model. In addition, the data used for the output results of the basic model in the training samples cannot be used for training, so as to prevent overfitting of the final prediction. Note that validation on the training set is better than that of on the test set. In order to prevent data leakage, it is necessary to output the results of each part of the sample separately by the K-Fold method. In our experiment, we use 5-Fold method (Figure 10): (1) we divide the data into five parts. One part at a time is used as the validation set, and the remaining four parts are used as the training set. In this way, a total of 5 models can be trained; (2) for the training set, one model is trained at a time to predict the validation set, and the prediction results are used as the second-layer input of the corresponding samples in the validation set. Repeat this process 5 times, we can obtain the outputs of each training sample that could be used as the input of the second-layer model; (3) for the test set, one model is trained at a time to predict a result. Therefore, the sample in the final test set will have 5 output results, and the average of these results will be used as the input for the second layer. Therefore, in our experiment, the following six machine learning algorithms, LIBSVM, DCA, LDA, KNN, RF and NB, are used to find the best combination of base-classifiers in the first stage and meta-classifier in the second stage. The three highest accurate classifiers, DAC, KNN and LDA, are used as the first layer. The training and test set are predicted based on the idea of 5-fold cross validation in each meta-model to prevent data leakage (Figure 10). Finally, the first layer of data input into the second layer of the DAC model, and this method has the highest accuracy.

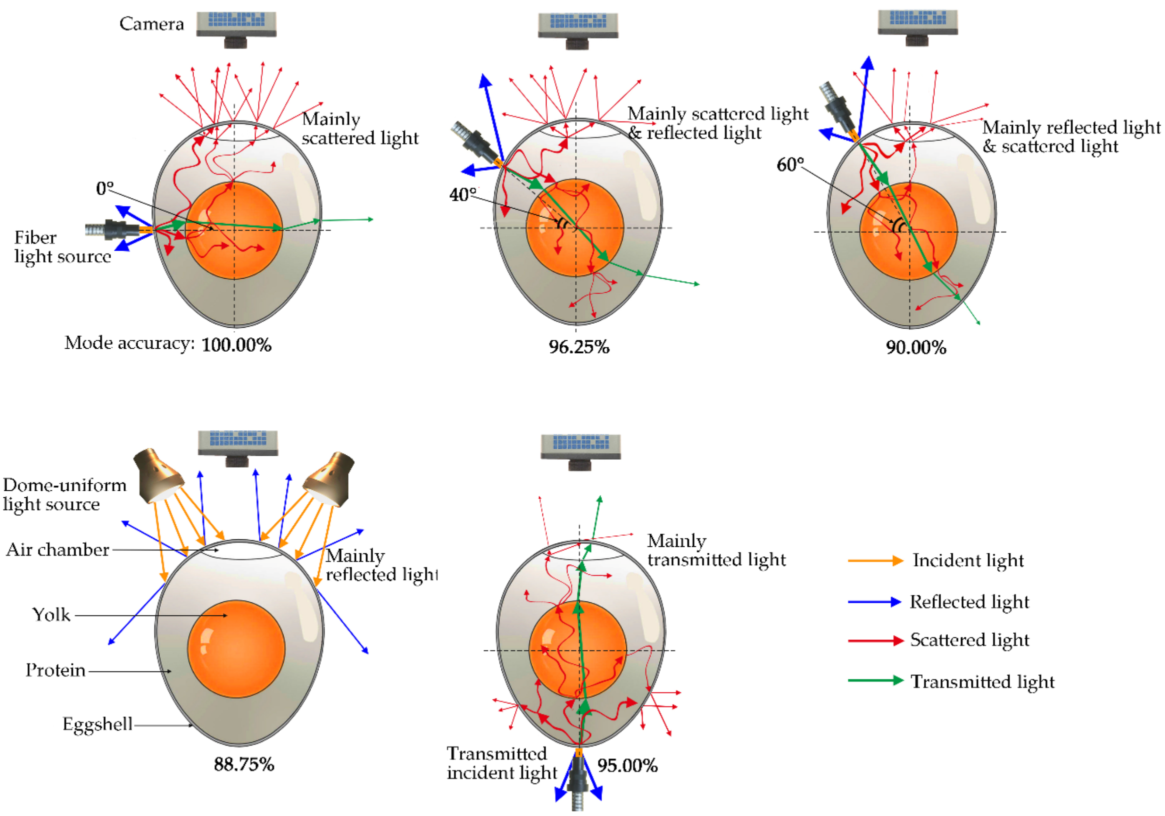


Figure 12. Light propagation inside an egg

Different incident angles cause different information contained in the light collected by the camera, resulting in different accuracy of egg freshness. The freshness is closely related to the internal composition of an egg, yolk index [38], the pH of protein [1] and air chamber index [39]. The spectra collected about the more internal information of an egg is the precondition for establishing a model with higher accuracy. The analysis of the light propagation paths inside an egg helps us understand the information contained in the image at different incident angles. For different incident mode, the propagation paths of light through an egg are different, so the information collected is also different (Figure 12). The camera mainly captures the reflected light of an egg as the incident light is a Dome-uniform light source, captures the scattered light through an egg as the incident angle is 0°, captures the reflection and the scattered light as the incident angle ranging from 0° and 60°, and captures the transmission light as the transmission fiber light. The scattered light through an egg carries out a lot of the biochemical information of the egg yolk, egg white and air chamber. The reflected light by an egg only contain the information of the eggshell. The transmission light through an egg also carries out a lot of information, and the camera will collect a higher proportion of the original light from the incident light source, resulting in a low accuracy. In our experiment, a camera captures a larger proportion of scattered light and a smaller proportion of reflection light as the incident angle is 0°, the accuracy of this angle is the highest. Meanwhile, the proportion of scattered light decreases and that of the reflection increase as the incident angle increase gradually from 0° to 60°, causing that their accuracies decease gradually with the increase of the incident angle. The proportion of the reflection should be the highest as the angle increase to 90°, thus its corresponding accuracy should be the lowest. In this mode, most of the lights are reflected by the eggshell and captured by a camera. A small part of the light passes through the egg shell to enter the inside of the egg, but a larger proportion of them shoot out from the bottom of the egg, which can not be detect by the camera on the top of the egg. Therefore, only a very small part of the light is scattered on the upper of the egg and captured by the camera, resulting in its low accuracy. However, the 90° incident angle could not

 doi:10.20944/preprints202008.0378.v1

be tested due to the location conflict of the camera and the incident light source. While, the dome-uniform light source is the light source with a weak intensity, which cannot nearly penetrate the egg shell and only reflection light could be captured by a camera. Thus, it is very similar to the 90°incident angle of fiber light source. This is the reason why the accuracy of the model decreases linearly as the angle increase from 0°to 60°and R. (Table 10). For the transmitted light source, most of the light is reflected from the bottom of the egg, the scattered light from the lower layer of the egg is absorbed by the yolk, and only a small part of the scattered light from the upper layer is captured by the camera, and a large amount of original light will also interfere with the test accuracy. Hence, its detection accuracy is not high.

5. Conclusions

This paper has studied a method for improving the accuracy of egg freshness based on scattering hyperspectral, researched the influence of different incident angles on the accuracy and explained its mechanism. The data processing process and conclusions are as the followings: (a) we established the classification model of egg freshness based on the combination of different preprocessing, feature wavelength extraction and weak classifiers, and obtained the best classification models. It is found that the 0° fiber light source-MSC-SPA-DAC has the highest accuracy of 96.25%. Moreover, the detection accuracy of 30° fiber light source-MA-CARS-KNN and 40° fiber light source-MSC-CARS-DAC are 95% and 95% respectively; (b) stacking ensemble learning is used to establish a fast egg freshness classification model to further improve the accuracy. In the 0° fiber optic light source-MSC-SPA-Stacking combination mode, the accuracy are increased from 96.25% to 100%; (c) the hyperspectral classifier model of egg freshness was established under different incident light irradiation. Their highest accuracy of scattering, reflection, transmission and mixed modes are 100.00%, 88.75%, 95.00% and 96.25%, respectively, indicating that the scattering hyperspectral for egg freshness detection is better than the other three. Moreover, the accuracy is inversely proportional to the incident angle, that is, the greater the incident angle, the lower the detection accuracy. This experiment finally realizes the non-destructive and high-precision detection of egg freshness based on scattering hyperspectral, and it has potential applications in online non-destructive detection.

Author Contributions: D. J. D. proposed the conceptualization and methodology, and wrote the paper. T. J. designed and carried out the experiments. W. L. improved the methodology and conceived the experiment. X. S. and R. X. programed the software. J. W. Z. compared the performance of the algorithms. All authors reviewed the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (31960487, 11604154), Jiangsu Provincial Natural Science Foundation of China (BK20181315), Yangzhou Key R&D Program (Modern Agriculture) (YZ2018038)

Acknowledgments: The authors wish to thank the editor and reviewers for their suggestions and thank Wei Lu for his guidance.

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

References

1. Dong, X.; Dong, J.; Li, Y.; Xu, H.; Tang, X. Maintaining the predictive abilities of egg freshness models on new variety based on VIS-NIR spectroscopy technique. *Computers and Electronics in Agriculture*, **2019**, *156*, *669–676*.

2. Liu, Y.; Ren, X.; Yu, H.; Cheng, Y.; Guo, Y.; Yao, W.; Xie, Y. Non-destructive and online egg freshness assessment from the egg shell based on Roman spectroscopy. *Food Control*, **2020**, *118*, *107426*.
3. Joshi, R.; Lohumi, S.; Joshi, R.; Kim, M. S.; Qin, J.; Baek, I.; Cho, B. Raman spectral analysis for non-invasive detection of external and internal parameters of fake eggs. *Sensors and Actuators B: Chemical*, **2019**, *303*, *127243*.
4. Lau, S.; Subbiah, J. An automatic system for measuring dielectric properties of foods: albumen, yolk, and shell of fresh eggs. *Journal of Food Engineering*, **2018**, *223*, *79–90*.
5. Sun, J.; Liu, B.; Mao, H.; Wu, X.; Gao, H.; Yang, N. Non-destructive examination for freshness of eggs based on dielectric properties and yolk index regression model. *Transactions of the Chinese Society of Agricultural Engineering*, **2016**, *32*, *290-295*.
6. Xiang, X.; Wang, Y.; Yu, Z.; Ma, M.; Zhu, Z.; Jin, Y. Non-destructive characterization of egg odor and fertilization status by SPME/GC-MS coupled with electronic nose. *Journal of the Science of Food and Agriculture*, **2019**, *99*, *3264–3275*.
7. Yimenu, S.; Kim, J.; Kim, B. Prediction of egg freshness during storage using electronic nose. *Poultry Science*, **2017**, 96, 3733–3746.
8. Sun, L.; Yuan, L.; Cai, J.; Lin, H.; Zhao, J. Egg freshness on-line estimation using machine vision and dynamic weighing. *Food Analytical Methods*, **2014**, *8*, *922–928*.
9. Guanjun, B.; Mimi, J.; Yi, X.; Shibo, C.; Qinghua, Y. Cracked egg recognition based on machine vision. *Computers and Electronics in Agriculture*, **2019**, *158*, *159–166*.
10. Suktanarak, S.; Teerachaichayut, S. Non-destructive quality assessment of hens' eggs using hyperspectral images. *Journal of Food Engineering*, **2017**, *215*, *97–103*.
11. Siche, R.; Vejarano, R.; Aredo, V.; Velasquez, L.; Saldaña, E.; Quevedo, R. Evaluation of food quality and safety with hyperspectral imaging (HSI). *Food Engineering Reviews*, **2015**, *8*, *306–322*.
12. Zhang, W.; Pan, L.; Tu, S.; Zhan, G.; Tu, K. Non-destructive internal quality assessment of eggs using a synthesis of hyperspectral imaging and multivariate analysis. *Journal of Food Engineering*, **2015**, *157*, *41–48*.
13. Lin, H.; Zhao, J.; Sun, L.; Chen, Q.; Zhou, F. Freshness measurement of eggs using near infrared (NIR) spectroscopy and multivariate data analysis. *Innovative Food Science & Emerging Technologies*, **2012**, *12*, *182–186*.
14. Huang, Y.; Liu, Y.; Yang, Y. Discrimination of tomato color grade based on spatial resolved spectroscopy and visible/near infrared spectroscopy. *Spectroscopy and Spectral Analysis*, **2019**, *39*, *3585-3591*.
15. Lu, Y.; Huang, Y.; Lu, R. Innovative Hyperspectral Imaging-Based Techniques for Quality Evaluation of Fruits and Vegetables: A Review. *Applied Sciences*, **2017**, *7*, *189*.
16. Qiu, Y.; Wu, G.; Xiao, Z.; Guo, Y.; Zhang, X.; Liu, K. An extreme-learning-machine-based hyperspectral detection method of insulator pollution degree. *IEEE Access*, **2019**, *7*, *121156–121164*.
17. Eisen, E.; Bohren, B..; McKean, H. The Haugh unit as a measure of egg albumen quality. *Poultry Science*, **1962**, *41*, *1461–1468*.
18. Zhao, J.; Fang, Y.; Chu, G.; Yan, H.; Hu, L.; Huang, L. Identification of leaf-scale wheat powdery mildew (blumeria graminis f. sp. tritici) combining hyperspectral imaging and an SVM classifier. *Plants*, **2020**, *9*, *936*.
19. Nelson, G.; Lines, A.; Bello, J.; Bryan, S. Online monitoring of solutions within microfluidic chips: simultaneous Raman and UV-vis absorption spectroscopies. *ACS Sensors*, **2019**, *4*, *2288-2295*.
20. Du, T.; Wen, G.; Cai, Z.; Zheng, W.; Tan, M.; Li, Y. Spectral clustering algorithm combining local covariance matrix with normalization. *Neural Computing and Applications*, **2018**, *32*, *6611–6618*.
21. Lohumi, S.; Lee, H.; Kim, M.; Qin, J.; Cho, B. Raman hyperspectral imaging and spectral similarity analysis for quantitative detection of multiple adulterants in wheat flour. *Biosystems Engineering*, **2019**, *181*, *103–113*.
22. Amodio, M.; Capotorto, I.; Chaudhry, M.; Colelli, G. The use of hyperspectral imaging to predict the distribution of internal constituents and to classify edible fennel heads based on the harvest time. *Computers and Electronics in Agriculture*, **2017**, *134*, *1–10*.
23. Qiu, X.; Jia X.; Zhao, H.; Zhang, C. Antinoise estimation of temperature and emissivity for FTIR spectrometer data using spectral polishing filters: design and comparison. *IEEE Transactions on Geoscience and Remote Sensing*, **2020**, *1-17*. *doi:10.1109/tgrs.2020.3008783*
24. Benes, E.; Fodor, M.; Kovács, S.; Gere, A. Application of detrended fluctuation analysis and yield stability index to evaluate near infrared spectra of green and roasted coffee samples. *Processes*, **2020**, *8*, *913*.

 doi:10.20944/preprints202008.0378.v1

25. Weng, H.; Lv, J.; Cen, H.; He, M.; Zeng, Y.; Hua, S.; He, Y. Hyperspectral reflectance imaging combined with carbohydrate metabolism analysis for diagnosis of citrus Huanglongbing in different seasons and cultivars. *Sensors and Actuators B: Chemical*, **2018**, *275*, *50–60*.
26. Hong, Y.; Liu, Y.; Chen, Y.; Liu, Y.; Yu, L.; Liu, Y.; Cheng, H. Application of fractional-order derivative in the quantitative estimation of soil organic matter content through visible and near-infrared spectroscopy. *Geoderma*, **2019**, *337*, *758–769*.
27. Li, Y.; Deng, L.; Yang, X.; Liu, Z.; Zhao, X.; Huang, F.; Zhu, S.; Chen, X.; Chen, Z.; Zhang, W. Early diagnosis of gastric cancer based on deep learning combined with the spectral-spatial classification method. *Biomedical Optics Express*, **2019**, *10*, *4999-5014*.
28. Wu, N.; Jiang, H.; Bao, Y.; Zhang, C.; Zhang, J; Song, W.; Liu, F. Practicability investigation of using near-infrared hyperspectral imaging to detect rice kernels infected with rice false smut in different conditions. *Sensors and Actuators B: Chemical*, **2020**, *308*, *127696*.
29. Medina, M.; Díaz, J.; Vignolo, C. Fractal dimension of sparkles in automotive metallic coatings by multispectral imaging measurements. *ACS Applied Materials & Interfaces*, **2014**, *6*, *11439–11447*.
30. Goudarzi, N.; Goodarzi, M.; Araujo, M.; Galvão, R. QSPR modeling of soil sorption coefficients (KOC) of pesticides using SPA-ANN and SPA-MLR. *Journal of Agricultural and Food Chemistry*, **2009**, *57*, *7153–7158*.
31. De Boves Harrington, P. support vector machine classification trees. *Analytical Chemistry*, **2015**, *87*, *11065–11071*.
32. Yan, S.; Yan, X. Using labeled autoencoder to supervise neural network combined with k-Nearest Neighbor for visual industrial process monitoring. *Industrial & Engineering Chemistry Research*, **2019**, *58*, *9952–9958*.
33. Xing, J.; Luo, K.; Wang, H.; Jin, T.; Fan, J. Novel sensitivity study for biomass directional devolatilization by random forest models. *Energy & Fuels*, **2020**, *34*, *8414-8423*.
34. Susič, N.; Žibrat, U.; Širca, S.; Strajnar, P.; Razinger, J.; Knapič, M.; Gerič Stare, B. Discrimination between abiotic and biotic drought stress in tomatoes using hyperspectral imaging. *Sensors and Actuators B: Chemical*, **2018**, *273*, *842–852*.
35. Sullivan, S.; Schmitt, P.; Muir, R.; DeWalt, E.; Simpson, G. Digital deconvolution filter derived from linear discriminant analysis and application for multiphoton Fluorescence Microscopy. *Analytical Chemistry*, **2014**, *86*, *3508–3516*.
36. Xin, Z.; Jun, S.; Xiaohong, W.; Bing, L.; Ning, Y.; Chunxia, D. Research on moldy tea feature classification based on WKNN algorithm and NIR hyperspectral imaging. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, **2019**, *206*, *378–383*.
37. Li, J.; Yao, Y.; Xu, H.; Hao, L.; Deng, Z.; Rajakumar, K. SecReT6: a web-based resource for type VI secretion systems found in bacteria. Environ. *Microbiol*, **2015**, *17*, *2196–2202*.
38. Wang, H.; Qiu, N.; Mine, Y.; Sun, H.; Meng, Y.; Li, B.; Keast, R. Quantitative Comparative Integrated Proteomic and Phosphoproteomic Analysis of Chicken Egg Yolk Proteins under Diverse Storage Temperatures. *Journal of Agricultural and Food Chemistry*, **2020**, *68*, *1057-1067*.
39. Li, Y.; Li, W.; Hu, L.; Svanberg, K.; Svanberg, S. Non-intrusive studies of gas contents and gas diffusion in hen eggs. *Biomedical Optics Express*, **2019**, *10*, *83-91*.