

## Article

# Application of Digital Image Analysis to the Prediction of Chlorophyll Content in *Astragalus* Seeds

Yanan Xu <sup>1</sup>, Keling Tu <sup>1</sup>, Ying Cheng <sup>1</sup>, Haonan Hou <sup>1</sup>, Hailu Cao <sup>2</sup>, Xuehui Dong <sup>1</sup> and Qun Sun <sup>1,\*</sup>

<sup>1</sup> College of Agronomy and Biotechnology, Department of Plant Genetics & Breeding and Seed Science, China Agricultural University, Beijing 100193, China; S20203010108@cau.edu.cn (Y.X.); b20193010029@cau.edu.cn (K.T.); chengying@cau.edu.cn (Y.C.); houhaonan@cau.edu.cn (H.H.); xuehuidong@cau.edu.cn (X.D.)

<sup>2</sup> Hengde Materia Medica (Beijing) Agricultural Technology Co., Ltd., Beijing 100070, China; caohailu@163.com

\* Correspondence: sunqun@cau.edu.cn; Tel.: +86-136-5123-4732

**Abstract:** Chlorophyll fluorescence (CF) has been applied to measure the chlorophyll content of seeds, in order to determine seed maturity, but the high price of equipment limits its wider application. *Astragalus* seeds were used to explore the applicability of digital image analysis technology to the prediction of seed chlorophyll content and to supply a low cost and alternative method. Our research comprised scanning and extracting the characteristic features of *Astragalus* seeds, determining the chlorophyll content, and establishing a predictive model of chlorophyll content in *Astragalus* seeds based on characteristic features. The results showed that the  $R^2$  of the MLR prediction model established with multiple features was  $\geq 0.947$ , and the  $R^2$  of the MLP model was  $\geq 0.943$ . By sorting of two single features, the R and G values, the  $R^2$  reached 0.969 and 0.965, respectively. A germination result showed that the lower the chlorophyll content, the higher the quality of the seeds. Therefore, we draw a conclusion that digital image analysis technology can be used to predict effectively the chlorophyll content of *Astragalus* seeds, and provide a reference for the selection of mature and viable *Astragalus* seeds.

**Keywords:** *Astragalus* seeds; chlorophyll content; multiple linear regression (MLR); multilayer perceptron (MLP); R; G

check for  
updates

**Citation:** Xu, Y.; Tu, K.; Cheng, Y.; Hou, H.; Cao, H.; Dong, X.; Sun, Q. Application of Digital Image Analysis to the Prediction of Chlorophyll Content in *Astragalus* Seeds. *Appl. Sci.* **2021**, *11*, 8744. <https://doi.org/10.3390/app11188744>

Academic Editor:  
Giuseppe Lacidogna

Received: 18 July 2021  
Accepted: 17 September 2021  
Published: 19 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Astragalus* is a large genus of plants in the botanical family Fabaceae, several species of which have been historically used in traditional Chinese medicine. As used here, “*Astragalus*” mainly refers to the species *Astragalus mongholicus* and *A. membranaceus*. *Astragalus* is one of the most popular medicinal herbs in China, and it has a long history of use as a medicinal material. *Astragalus* is one of 40 commonly used bulk medicinal materials, and the demand for it is increasing daily [1,2]. However, the quality of traditional Chinese medicinal plant seeds is far behind that of crop seeds, and there are many problems, such as mixed provenance, low purity, inconsistent maturity, low germination rate, and slow and irregular emergence. Seeds have become the weakest link in the production of traditional Chinese medicinal plants and are the “bottleneck” that restricts the standardized production and development of traditional Chinese medicine.

Seed maturity is closely related to seed vigor [3,4]. The comprehensive features such as bud rate, bud potential, soil arching ability, seedling rate, and seedling uniformity of seeds with poor maturity decrease, which seriously affects the quality of seeds and the entire production system. Therefore, it is necessary to improve seed maturity. Chlorophyll content in the seeds of many species can be used as an indicator of seed maturity and is a major contributor to seed quality [5–8].

Using chlorophyll fluorescence (CF) to determine seed maturity is a relatively new technology. CF seed detectors are used to detect the chlorophyll fluorescence (blue) gen-

erated by pulsed light. Fluorescence emissions at wavelengths of 600 nm and above from all fluorescent substances will be detected. The pulsed light emitted by the detector can separate the fluorescence from the surrounding light. Therefore, the detector can calculate the rough chlorophyll concentration by measuring chlorophyll fluorescence. Keenanoglu et al. [9] classified the different stages of maturity in four varieties of pepper seeds based on CF. After classification, the germination percentage and vigor of the seeds with low CF were shown to be the highest. Jalink et al. [10,11] used the magnitude of the CF to judge and grade seed maturity in cabbage. These authors found that the smaller the signal value, the higher the seed germination rate. Although CF is widely used and is highly portable, it still has the problem of high cost. Digital image analysis technology is a low-cost and non-destructive technology. Images can be obtained by classical reflection technologies, such as optical and micrography methods, or by more modern technologies, such as thermal imaging, fluorescence and magnetic resonance [12–16]. This technology can obtain seed image information, extract the shape, color, and texture information of seeds, and is widely used in seed color feature determination [17,18], the improvement of seed vigor and quality monitoring [19–22], crop seed quality classification [23–25], variety identification [26,27], purity determination [28], quantifying intraspecific genetic diversity [29] and evaluating crop and seed maturity [30–32]. Digital image analysis technology is a relatively mature technology, and its application to the prediction of chlorophyll content in *Astragalus* seeds is highly feasible with excellent development prospects.

Therefore, it is important to develop a fast, non-destructive, low-cost and efficient method to measure the chlorophyll content of *Astragalus* seeds. In this study, we determined the chlorophyll content of *Astragalus* seeds and the characteristic features of the seeds were extracted. The models were established by using the characteristic features as the input factor of the model, the chlorophyll content as the output parameter, and the multiple linear regression (MLR) and multi-layer perceptron (MLP) methods to establish the model. A single-feature prediction model using the R and G values was attempted to provide a new idea for the prediction of chlorophyll content in *Astragalus* seeds.

## 2. Materials and Methods

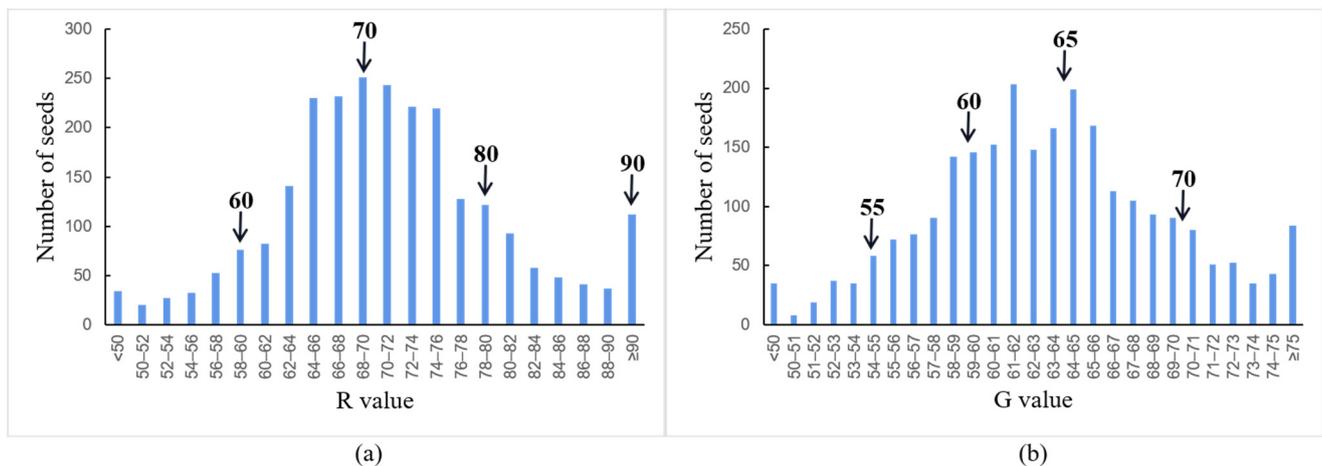
### 2.1. Materials

Seeds of *A. mongholicus* were collected from the Anguo market in Hebei Province in 2020.

### 2.2. Seed Scanning and Extraction of Phenotypic Features

The *Astragalus* seeds were divided into two groups based on color: black-brown and yellow. The seeds of the two groups were then mixed in a certain proportion, and there were a total of 50 groups with 5 g of seeds in each group. A Scanmaker i360/i460 scanner was used for group scanning at a resolution of 300 dpi. The images were saved in TIFF lossless format. The Phenoseed automatic extraction system, which was jointly developed by the Seed Science and Technology Research Center of China Agricultural University (Beijing, China) and Nanjing Zhinong Yunxin Big Data Technology Co., Ltd. (Nanjing, China), was used to extract seed phenotypic features. The color features included R (red in the primary color), G (green in the three primary color light), B (blue in the primary color light), L (luminosity), a (range from red to green), b (range from blue to yellow), hue, saturation, value, gray, and standard deviation, for a total of 20 color features. The average seed color feature of each group was taken.

In order to explore the relationship between R, G values and chlorophyll content, 5000 seeds of *Astragalus* were selected for single seed scanning and seed phenotype feature extraction. The frequency histogram of R and G values of 5000 *Astragalus* seeds was shown in Figure 1. We divided 2500 seeds into five groups based on their R values:  $R \leq 60$ ,  $60 < R \leq 70$ ,  $70 < R \leq 80$ ,  $80 < R \leq 90$ ,  $R > 90$ . The remaining 2500 seeds were divided into five groups based on their G values:  $G \leq 55$ ,  $55 < G \leq 60$ ,  $60 < G \leq 65$ ,  $65 < G \leq 70$ ,  $G > 70$ . Each group of seeds weighed 2 g.



**Figure 1.** A frequency histogram of the R value (a) and G value (b) of *Astragalus* seeds.

### 2.3. Determination of Chlorophyll Content

The chlorophyll content of *A. mongholicus* seeds in each group was determined by a FluoMini Pro optical chlorophyll fluorescence (CF) detector provided by Beijing Youkede Agricultural Technology Co., Ltd. (Beijing, China), six times for each group, and the average value was taken.

### 2.4. Standard Germination Test

Five groups of seeds were divided into two groups based on the R and G values. The seeds were germinated on paper at 25 °C with 24 h illumination. There were 100 seeds per replicate and three replicates per group. The seeds were observed for seven days, and moldy or rotten seeds were removed. On the 7th day of germination, the germination percentage, mildew percentage, number of abnormal seeds, percentage of dead seeds, percentage of hard seeds, and the percentage of high-quality seeds were recorded. Many studies have shown that the quality of hard seeds is better than that of non-hard seeds [33–35]. Therefore, in this study we used the following formula:

$$\text{High-quality seed percentage (\%)} = \text{Germination percentage (\%)} + \text{hard seed percentage (\%)}$$

### 2.5. Data Analysis

All data were processed and analyzed using Microsoft Office Excel 2019 and IBM SPSS statistics 23.

All color features, principal component factors extracted from all color features, all color average features, principal component factors extracted from all color average features, R\_mean, G\_mean, B\_mean, and Gray\_mean were independent variables, chlorophyll content was the dependent variable, and the MLR and MLP models were established.

A correlation analysis between each of the color features and chlorophyll content was performed. Two features, R\_mean and G\_mean, were used to predict the chlorophyll content of *Astragalus* seeds.

#### 2.5.1. Principal Component Analysis (PCA)

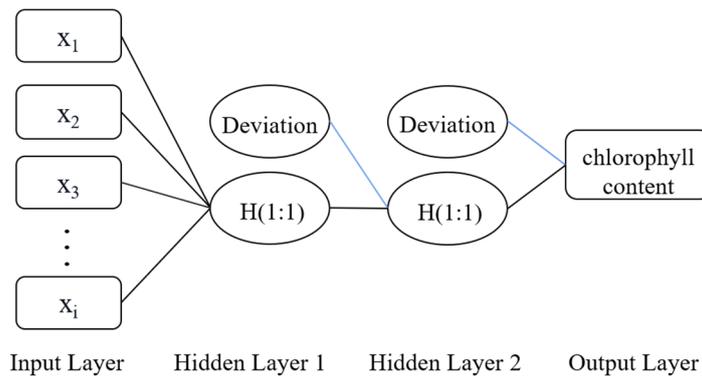
PCA is a linear dimension reduction method, which simplifies multiple indicators into a small number of comprehensive indicators and uses a small number of variables to reflect the information of the original variables as much as possible to ensure that the loss of original information is small and the number of variables is as small as possible [36].

In this study, we selected the maximum variance method of the rotation method and the regression method for the factor scores. PCA was performed on all color features and all color average features to simplify the model.

### 2.5.2. Multilayer Perceptron Network (MLP)

MLP is one of the most common neural network topologies. The structure of the MLP is basically similar to a set of cascade perceptrons, which maps a group of input vectors to a group of output vectors. The input and output can be connected by multi-layer weighting, and has strong self-learning, adaptive, associative memory, parallel processing, and other capabilities for things and environment [37,38].

Figure 2 shows the MLP topology used in this study. A MLP network with two hidden layers was selected for this study. The hidden layer used the hyperbolic tangent activation function of SPSS. The output layer used the softmax activation function of SPSS. Training set: the test set was 7:3; that is, there were 35 samples in the training set and 15 samples in the test set.



**Figure 2.** Multilayer perceptron network (MLP) topology used in this research.

### 2.5.3. Multiple Linear Regression (MLR)

In regression analysis, multiple regression involves two or more independent variables. In fact, a single phenomenon is often associated with multiple factors. It is more effective and practical to predict or estimate the dependent variable by the optimal combination of multiple independent variables than by using only one independent variable. In this study, the step-by-step method in SPSS was used to establish the MLR chlorophyll content prediction model. The setting of the model sample set was the same as in Section 2.5.2.

The specific experimental process using in this study is shown in Figure 3.

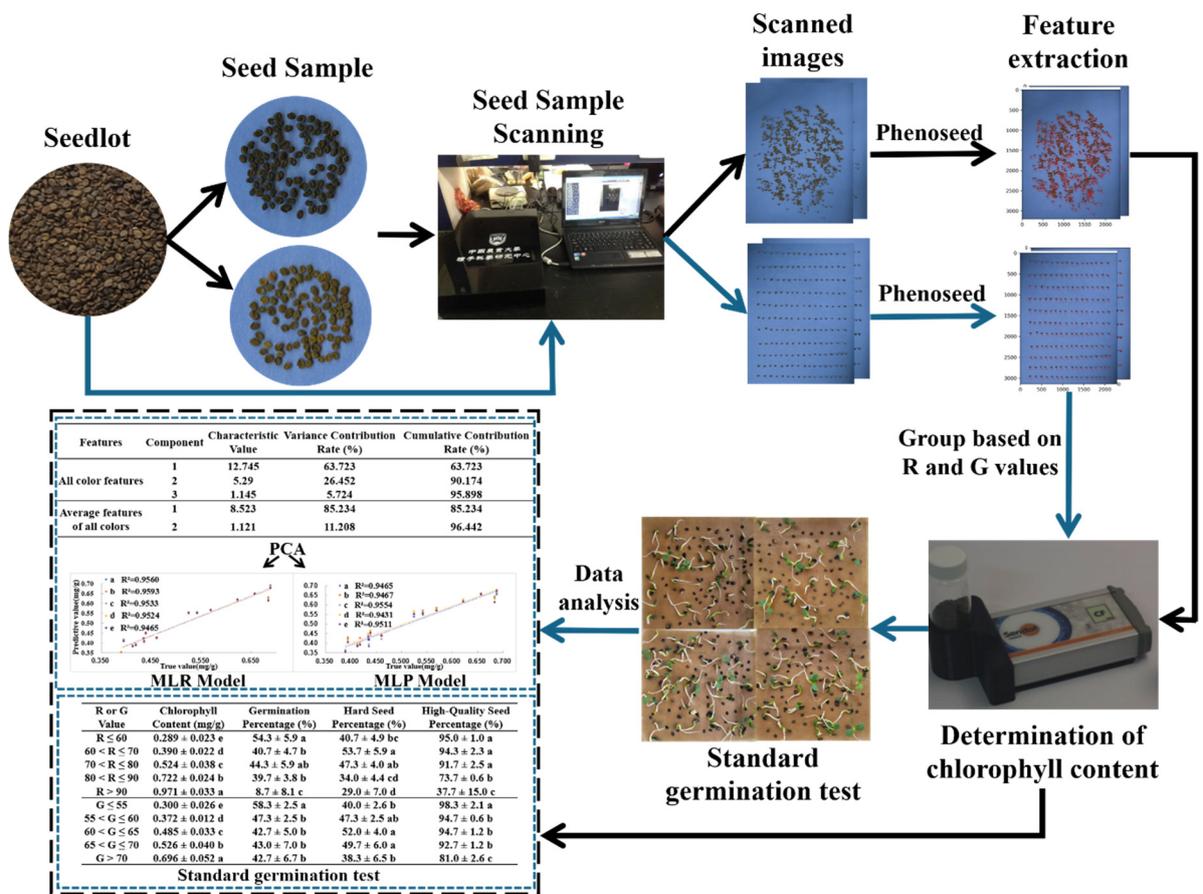


Figure 3. Flowchart of the main procedures for extracting and analyzing the data.

### 3. Results

#### 3.1. Establishment of the Chlorophyll Content Prediction Model

##### 3.1.1. Principal Component Analysis (PCA)

All color features and all color average features were analyzed by principal component analysis, as shown in Table 1. Three principal component factors were extracted from all color features, and the cumulative variance contribution rate was 95.898%, indicating that the extracted factors can reflect  $\geq 95\%$  of the original information for all variables. Two principal component factors were finally extracted from all the color average features, and the cumulative variance contribution rate was 96.442%, which indicated that the extracted factors could reflect  $\geq 96\%$  of the original information for all variables.

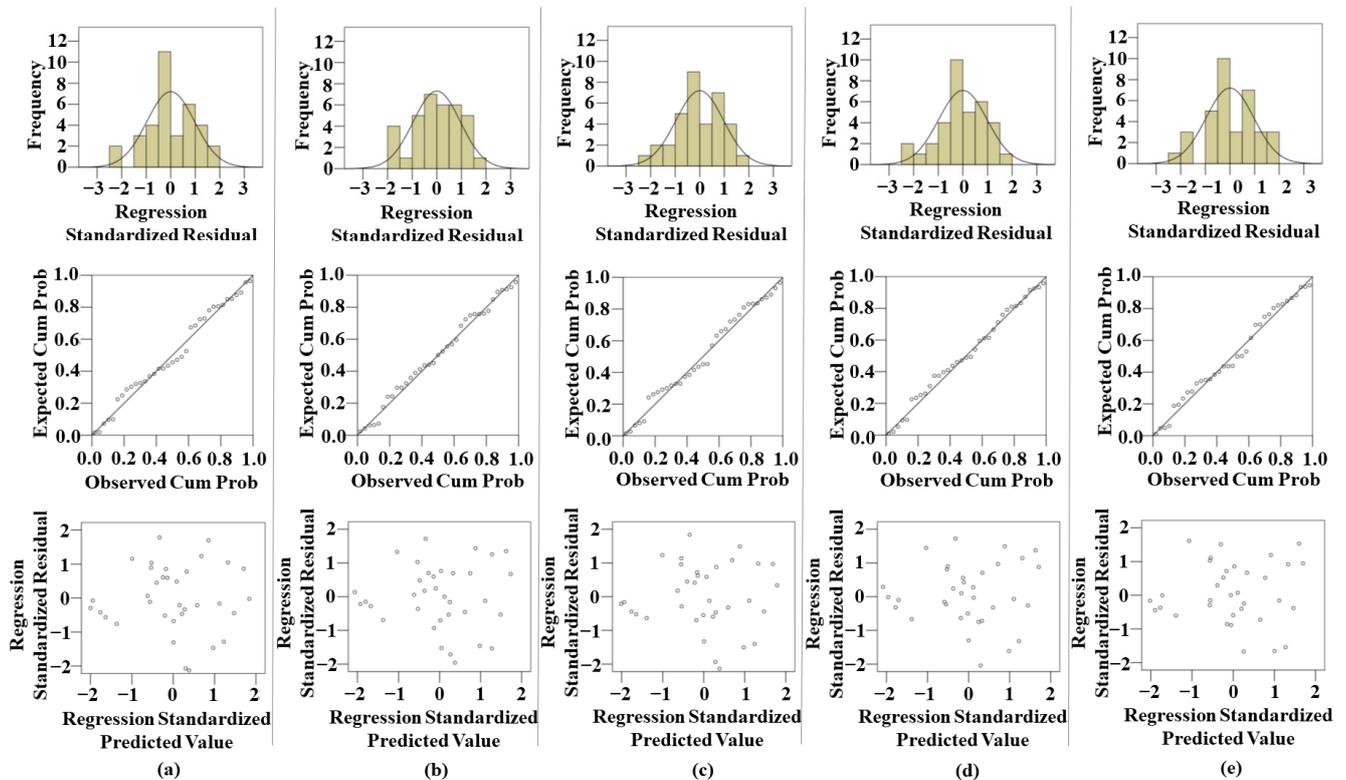
Table 1. The principal component eigenvalues, contribution rate, and cumulative contribution rate of all color features and all color average features.

| Features                       | Component | Characteristic Value | Variance Contribution Rate (%) | Cumulative Contribution Rate (%) |
|--------------------------------|-----------|----------------------|--------------------------------|----------------------------------|
| All color features             | 1         | 12.745               | 63.723                         | 63.723                           |
|                                | 2         | 5.29                 | 26.452                         | 90.174                           |
|                                | 3         | 1.145                | 5.724                          | 95.898                           |
| Average features of all colors | 1         | 8.523                | 85.234                         | 85.234                           |
|                                | 2         | 1.121                | 11.208                         | 96.442                           |

### 3.1.2. Establishment and Verification of the MLR Model

We used all color features (a), principal component factors extracted from all color features (b), all color average features (c), principal component factors extracted from all color average features (d), R\_mean, G\_mean, B\_mean, and Gray\_mean (e) as input factors, and chlorophyll content as the output parameter. Five MLR models of chlorophyll content in *Astragalus* seeds were established by stepwise regression.

The residual analysis results are shown in Figure 4. It can be seen from the histogram of standardized residuals that the standardized residuals of the five models basically obey the normal distribution with a mean value of 0 and a standard deviation of 1. At the same time, it can be seen from the normal probability diagram (P-P diagram) that the scatter basically walks around the diagonal of the first quadrant, so it is judged that the residual basically obeys the normal distribution. It can be seen from the scatter diagram of standardized residuals and standardized predicted values that the scatter fluctuation range of standardized residuals basically remained stable and did not change with the change of standardized predicted values, so it met the homogeneity of variance.



**Figure 4.** Residual analysis results of five multiple linear regression (MLR) models. (a) All color features; (b) principal component features extracted from all color features by principal component analysis; (c) average features of all colors; (d) principal component features extracted from principal component analysis of color average features; (e) R\_mean, G\_mean, B\_mean, and Gray\_mean.

Generally, the adjusted  $R^2$  represents the degree that the linear equation can reflect the real data. The Durbin–Watson statistic is usually used to determine whether there is autocorrelation between data. The variance inflation factor (VIF) value is usually used to judge whether there is collinearity between independent variables. If the VIF is  $\leq 10$ , it means that there is no collinearity between independent variables, and the model operation is more accurate.

The prediction results of the model are shown in Table 2. It can be seen from Table 2 that there were no significant differences among the five models, and the adjusted  $R^2$  was about 0.98, which indicates that the linear equation reflects the real data better. In this study, the

Durbin–Watson statistics were around 1.8, close to 2, and the  $p$ -values were  $>0.05$ , which basically shows that there was no autocorrelation in the data.

**Table 2.** Summary of multiple linear regression (MLR) prediction models for chlorophyll content of *Astragalus* seeds.

| Models | R     | R <sup>2</sup> | Adjusted R <sup>2</sup> | Root-Mean-Square Error (RMSE) | Durbin–Watson   |           |       |
|--------|-------|----------------|-------------------------|-------------------------------|-----------------|-----------|-------|
|        |       |                |                         |                               | Autocorrelation | Statistic | $p$   |
| a      | 0.991 | 0.981          | 0.980                   | 0.014                         | 0.147           | 1.706     | 0.316 |
| b      | 0.990 | 0.979          | 0.977                   | 0.015                         | 0.048           | 1.886     | 0.628 |
| c      | 0.990 | 0.981          | 0.980                   | 0.014                         | 0.091           | 1.814     | 0.495 |
| d      | 0.991 | 0.982          | 0.981                   | 0.013                         | 0.098           | 1.793     | 0.451 |
| e      | 0.990 | 0.980          | 0.979                   | 0.014                         | 0.099           | 1.769     | 0.409 |

a: All color features; b: Principal component factors extracted from all color features by principal component analysis; c: Average features of all colors; d: Principal component features extracted from principal component analysis of color average features; e: R\_mean, G\_mean, B\_mean, and Gray\_mean.

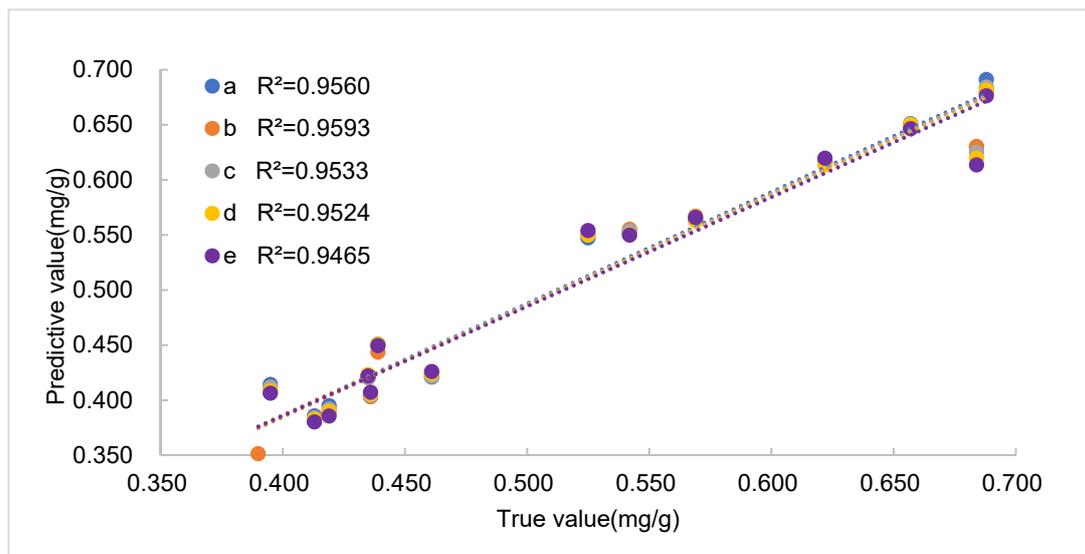
Table 3 shows the prediction results of the five MLR models. It can be seen that the corresponding significance levels of the independent variables in each model were  $\leq 0.05$ , indicating that the above independent variables could have a significant impact on the chlorophyll content of the dependent variable. In addition, the VIF indicates that there was no collinearity between the independent variables, and the model operation was more accurate.

**Table 3.** Results of multiple linear regression (MLR) prediction models for chlorophyll content of *Astragalus* seeds.

| Models | Variables  | Coefficient | Standard Deviation | Standardization Coefficient | Sig.  | Variance Inflation Factor (VIF) |
|--------|------------|-------------|--------------------|-----------------------------|-------|---------------------------------|
| a      | (constant) | −1.166      | 0.154              |                             | 0.000 |                                 |
|        | R_mean     | 0.018       | 0.000              | 1.026                       | 0.000 | 1.274                           |
|        | a_std      | 0.229       | 0.076              | 0.082                       | 0.005 | 1.274                           |
| b      | (constant) | 0.502       | 0.002              |                             |       |                                 |
|        | Factor 1   | 0.088       | 0.002              | 0.960                       | 0.000 | 1.042                           |
|        | Factor 2   | −0.017      | 0.002              | −0.185                      | 0.000 | 1.003                           |
|        | Factor 3   | 0.044       | 0.003              | 0.418                       | 0.000 | 1.045                           |
| c      | (constant) | −0.961      | 0.093              |                             | 0.000 |                                 |
|        | R_mean     | 0.017       | 0.001              | 0.939                       | 0.000 | 1.504                           |
|        | a_mean     | 0.140       | 0.049              | 0.085                       | 0.008 | 1.504                           |
| d      | (constant) | 0.502       | 0.002              |                             | 0.000 |                                 |
|        | Factor 1   | 0.079       | 0.002              | 0.757                       | 0.000 | 1.006                           |
|        | Factor 2   | 0.061       | 0.002              | 0.701                       | 0.000 | 1.006                           |
| e      | (constant) | −0.544      | 0.078              |                             | 0.000 |                                 |
|        | R_mean     | 0.031       | 0.006              | 1.732                       | 0.000 | 158.208                         |
|        | G_mean     | −0.017      | 0.007              | −0.746                      | 0.025 | 158.208                         |

a: All color features; b: Principal component features extracted from all color features by principal component analysis; c: Average features of all colors; d: Principal component features extracted from principal component analysis of color average features; e: R\_mean, G\_mean, B\_mean, and Gray\_mean.

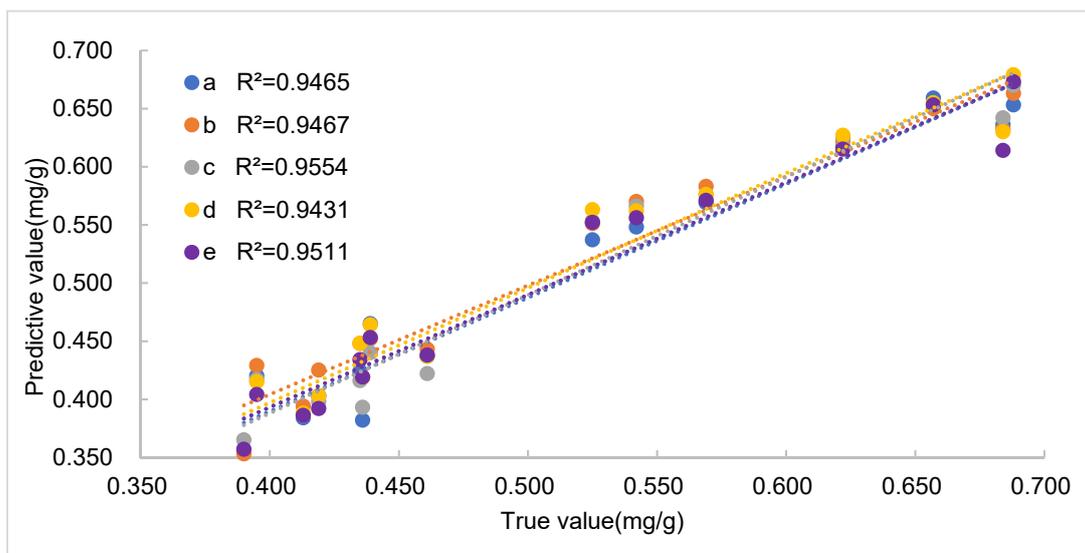
Fifteen samples of *Astragalus* seeds (the test set) were then used to verify the above five models, and the results are shown in Figure 5. It can be seen that the predicted values for chlorophyll content were very near the true values, the R<sup>2</sup> was above 0.947 in all cases, and the predicted results were better.



**Figure 5.** Five multiple linear regression (MLR) models with different characteristic features were established to predict the results. a: All color features; b: principal component features extracted from all color features by principal component analysis; c: average features of all colors; d: principal component features extracted from principal component analysis of color average features; e: R\_mean, G\_mean, B\_mean, and Gray\_mean.

### 3.1.3. Establishment and Verification of MLP Model

Similarly, using all color features (a), principal component factors extracted from all color features (b), all color average features (c), principal component factors extracted from all color average features (d), and R\_mean, G\_mean, B\_mean, and Gray\_mean (e) as input factors and chlorophyll content as the output parameter, five MLP models of chlorophyll content in *Astragalus* seeds were established. The  $R^2$  value predicted by the model is  $\geq 0.968$ . The test set of 15 samples of *Astragalus* seeds was used to verify the above five models, and the results are shown in Figure 6. It can be seen that the predicted values for chlorophyll content were near the true values, and the  $R^2$  values were  $\geq 0.943$ .



**Figure 6.** Multilayer perceptron (MLP) models with different characteristic features were established to predict the results. a: All color features; b: Principal component factors extracted from all color features by principal component analysis; c: Average features of all colors; d: Principal component factors extracted from principal component analysis of color average features; e: R\_mean, G\_mean, B\_mean, and Gray\_mean.

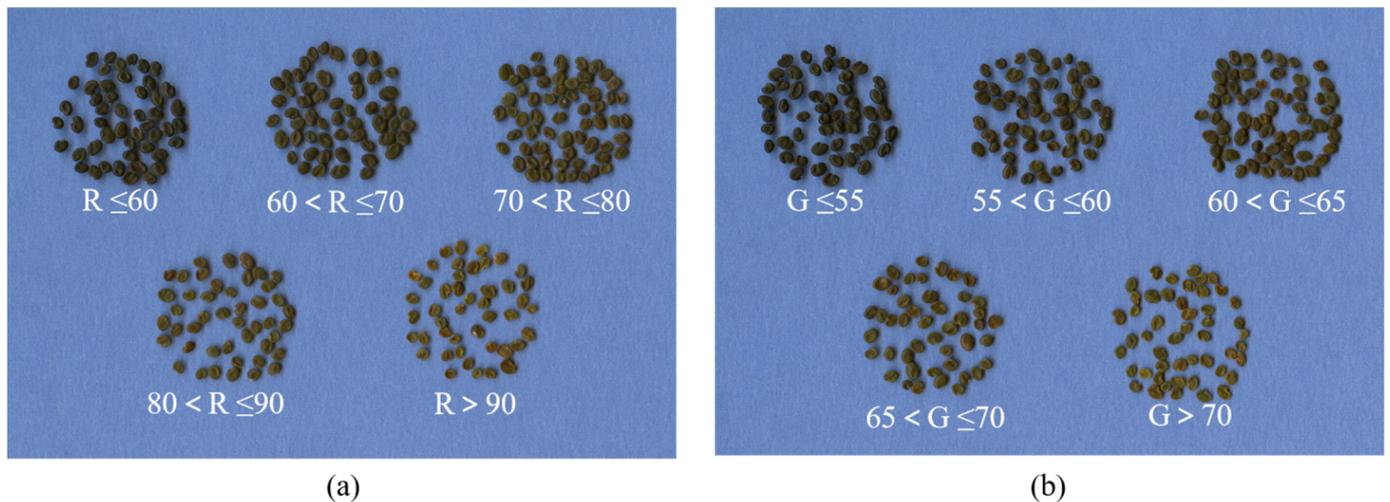
### 3.2. Prediction of Chlorophyll Content Using the R Value and G Value

It can be seen in Section 3.1.2 that the MLR model was established by the stepwise regression method for different characteristic features, and R\_mean was included in the model. The Pearson correlation coefficient between R\_mean and chlorophyll content was 0.980. Therefore, this study will continue to explore the relationship between the R value and chlorophyll content in order to achieve the purpose of using a single R\_mean to predict chlorophyll content. In general, we think that G value may be more related to chlorophyll content, so in order to explore more possibilities, we continued to examine the relationship between the G\_mean value and chlorophyll content, and to study the effect of G value on chlorophyll content prediction.

#### 3.2.1. R and G Value Distribution

The frequency histogram of R and G values of *Astragalus* seed batch showed that the R values of most seeds were between 60–90 and the G values of most seeds were between 55–70, but the R and G values of some seeds were not within this range (Figure 1). This showed that the R and G values of seed batches were not uniform, which may be due to the difference in chlorophyll content between different seeds.

Based on the distribution of R and G values, the seed batch was divided into five groups. The boundaries of the five sub samples were set as R values of 60, 70, 80 and 90 respectively; The G values were 55, 60, 65 and 70, respectively. The color difference between the seeds of the five subsamples was visible to the naked eye (Figure 7). The higher the R and G value, the lighter the seed coat color. The color difference features that chlorophyll may be located outside the seed coat.



**Figure 7.** *Astragalus* seeds from the R value (a) and G value (b) subsamples showing variations in color.

#### 3.2.2. Prediction of Chlorophyll Content Using the R Value and G Value

The results showed that there was a linear relationship between R value and the chlorophyll content of *Astragalus* seeds;  $R^2$  was 0.969, so the R value could be used to predict chlorophyll content (Figure 8a). The results showed that there was a linear relationship between G value and the chlorophyll content of *Astragalus* seeds; the  $R^2$  was 0.965, and the prediction result was still ideal (Figure 8b).

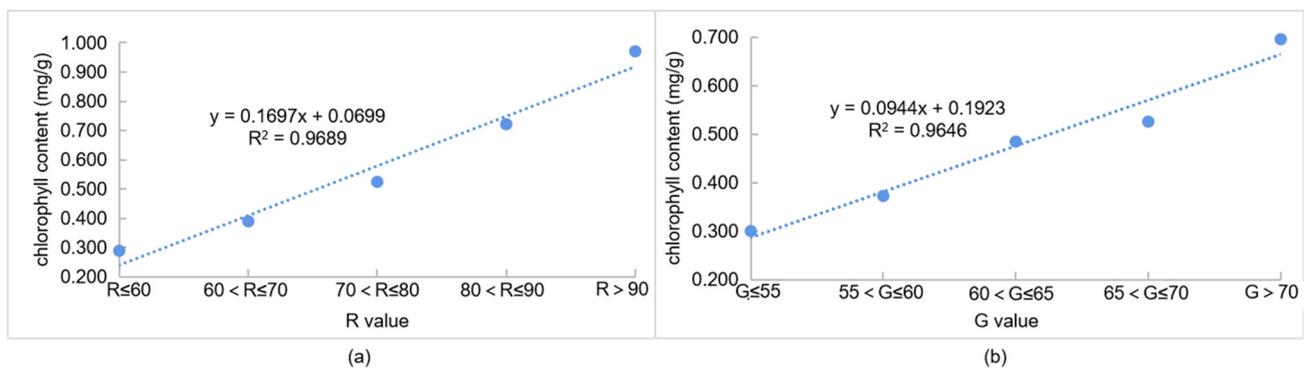


Figure 8. The linear relationships between the R value (a) and G value (b) and chlorophyll content.

### 3.3. Relationship between Seed Vigor and Chlorophyll Content of Astragalus Seeds

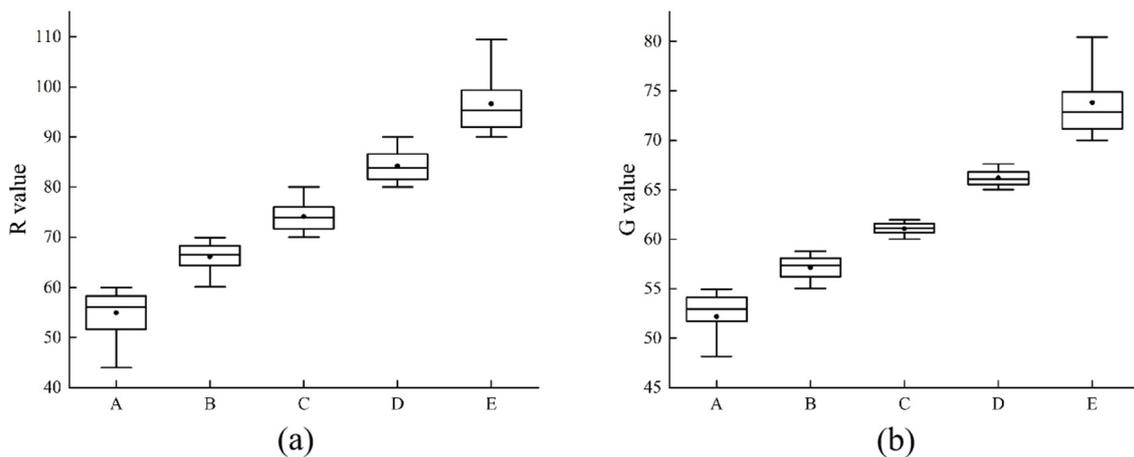
The chlorophyll contents of different *Astragalus* seeds could be predicted based on the R and G values. In order to further demonstrate the relationship between the vitality of *Astragalus* seeds and the chlorophyll content, we conducted germination tests of *Astragalus* seeds with known chlorophyll content and grouped them based on the R values and G values. The results are given in Table 4. The proportion of high-quality seeds of seeds with  $R > 80$  was significantly lower than that of seeds with  $R \leq 80$ . The germination percentage, hard-seed percentage, high-quality seed percentage, abnormal seedling percentage and mildew percentage of seeds with  $R > 90$  were significantly lower than those with  $R \leq 90$ , in which the germination percentage was only about 8.7% and the high-quality seed percentage was only about 37.7%. Similarly, G values show similar laws. The germination percentage and the percentage of the high-quality seed with  $G \leq 55$  were significantly higher than those with  $G > 55$ , and the percentage of the high-quality seed reached about 98.3%. Among the seeds with  $G > 55$ , the seeds with  $G > 70$  were significantly lower than other seeds in all aspects, and the percentage of the high-quality seed was about 81.0%.

Table 4. Statistics of germination percentage and hard seed percentage of *Astragalus* seeds in different R and G value groups.

| R or G Value     | Chlorophyll Content (mg/g) | Germination Percentage (%) | Hard Seed Percentage (%) | High-Quality Seed Percentage (%) |
|------------------|----------------------------|----------------------------|--------------------------|----------------------------------|
| $R \leq 60$      | $0.289 \pm 0.023$ e        | $54.3 \pm 5.9$ a           | $40.7 \pm 4.9$ bc        | $95.0 \pm 1.0$ a                 |
| $60 < R \leq 70$ | $0.390 \pm 0.022$ d        | $40.7 \pm 4.7$ b           | $53.7 \pm 5.9$ a         | $94.3 \pm 2.3$ a                 |
| $70 < R \leq 80$ | $0.524 \pm 0.038$ c        | $44.3 \pm 5.9$ ab          | $47.3 \pm 4.0$ ab        | $91.7 \pm 2.5$ a                 |
| $80 < R \leq 90$ | $0.722 \pm 0.024$ b        | $39.7 \pm 3.8$ b           | $34.0 \pm 4.4$ cd        | $73.7 \pm 0.6$ b                 |
| $R > 90$         | $0.971 \pm 0.033$ a        | $8.7 \pm 8.1$ c            | $29.0 \pm 7.0$ d         | $37.7 \pm 15.0$ c                |
| $G \leq 55$      | $0.300 \pm 0.026$ e        | $58.3 \pm 2.5$ a           | $40.0 \pm 2.6$ b         | $98.3 \pm 2.1$ a                 |
| $55 < G \leq 60$ | $0.372 \pm 0.012$ d        | $47.3 \pm 2.5$ b           | $47.3 \pm 2.5$ ab        | $94.7 \pm 0.6$ b                 |
| $60 < G \leq 65$ | $0.485 \pm 0.033$ c        | $42.7 \pm 5.0$ b           | $52.0 \pm 4.0$ a         | $94.7 \pm 1.2$ b                 |
| $65 < G \leq 70$ | $0.526 \pm 0.040$ b        | $43.0 \pm 7.0$ b           | $49.7 \pm 6.0$ a         | $92.7 \pm 1.2$ b                 |
| $G > 70$         | $0.696 \pm 0.052$ a        | $42.7 \pm 6.7$ b           | $38.3 \pm 6.5$ b         | $81.0 \pm 2.6$ c                 |

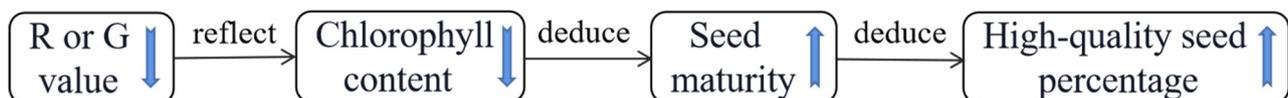
Notes: The means followed by different lowercase letters are significantly different at the 0.05 level.

Figure 9 is a box plot between R values (a) and G values (b) of different groups and their corresponding high-quality seed percentage. As can be seen from Figure 9, different high-quality seed percentage correspond to different R value and G value ranges.



**Figure 9.** Box plot between R values (a) and G values (b) of different groups and their corresponding high-quality seed percentage. (a) A: A seed population with  $95.0 \pm 1.0\%$  high-quality seed percentage; (a) B: A seed population with  $94.3 \pm 2.3\%$  high-quality seed percentage; (a) C: A seed population with  $91.7 \pm 2.5\%$  high-quality seed percentage; (a) D: A seed population with  $73.7 \pm 0.6\%$  high-quality seed percentage; (a) E: A seed population with  $37.7 \pm 15.0\%$  high-quality seed percentage; (b) A: A seed population with  $98.3 \pm 2.1\%$  high-quality seed percentage; (b) B: A seed population with  $94.7 \pm 0.6\%$  high-quality seed percentage; (b) C: A seed population with  $94.7 \pm 1.2\%$  high-quality seed percentage; (b) D: A seed population with  $92.7 \pm 1.2\%$  high-quality seed percentage; (b) E: A seed population with  $81.0 \pm 2.6\%$  high-quality seed percentage.

The lower the R and G values, the higher the percentage of high-quality seeds; that is, the lower the chlorophyll content, the higher the percentage of high-quality seeds (Figure 10). At the same time, this study also provided a reference for the selection of mature, viable *Astragalus* seeds.



**Figure 10.** The relation between the maturity and G or R values.

#### 4. Discussion

Seed maturity plays an important role in the production, processing, and marketing of *Astragalus* seeds. Chlorophyll content is not only related to photosynthesis, but also has an effect on crop yield. In seeds, the higher chlorophyll content, the lower the maturity. Stated another way, seed maturity is relatively high when the chlorophyll content is low [39].

The results of our study show that digital image analysis technology can be applied to the rapid and non-destructive prediction of chlorophyll content in *Astragalus* seeds. In this study, MLR and MLP models were established for the rapid prediction of chlorophyll content and we found that both the models were very effective in predicting the amount of chlorophyll present in *Astragalus* seeds. The  $R^2$  of the MLR model test set was  $\geq 0.947$ , and the  $R^2$  of the MLP model test set was  $\geq 0.943$ . There were no significant differences between the five types of model and the two modeling methods. Therefore, in practical application, only the R\_mean, G\_mean, B\_mean, and Gray\_mean are used to establish the model, which makes the sample imaging more convenient and faster.

We then used another batch of mature seeds of *Astragalus* to verify the established model. The predicted  $R^2$  of the training set was  $\sim 0.7$ , and the effect was not ideal. The high maturity batches of *Astragalus* seeds were used to establish the model, and the training set  $R^2$  was  $\sim 0.7$ . Two batches of *Astragalus* seeds were mixed and modeled, and the two batches were verified independently. The results showed that the predicted  $R^2$  of immature *Astragalus* seeds was  $\sim 0.9$ , while the predicted  $R^2$  of highly mature *Astragalus*

seeds was only ~0.6. This indicates that the model established in this study is suitable for the prediction of chlorophyll content in immature *Astragalus* seeds, but the prediction results for highly mature *Astragalus* seeds are not ideal, which merits further discussion.

We also found that the MLR model was established by the stepwise regression method through different characteristic features, and R\_mean was included in the model. The results showed that there is a linear relationship between R value and chlorophyll content; the  $R^2$  was 0.969, so the R value could be used to predict chlorophyll content. At the same time, our results also explored the relationship between the G value and chlorophyll content. These results showed that there is also a linear relationship between G value and chlorophyll content in *Astragalus* seeds. The  $R^2$  was 0.965, and the prediction result was worse than the R value. The reason for this may be that the chlorophyll content of the seeds is mainly reflected in the differences in the G values, but the similarity of G values in the seed groups was very high, which leads to a concentration of the G value distribution and a dense sorting interval. Based on this, the R value was found to be ideal. Therefore, in practical application, using the R value to predict chlorophyll content will give better results. At the same time, another batch of mature *Astragalus* seeds was scanned to extract the R value of each seed. The results showed that the R values of mature *Astragalus* seeds were more concentrated and the R values were lower, which also confirmed the previous results.

In order to further verify the relationship between seed vigor and chlorophyll content in *Astragalus*, we performed germination tests of *Astragalus* seeds with known chlorophyll contents that were grouped based on the R and G values. The results showed that the lower the R and G values, the lower the chlorophyll contents, and the higher the percentage of high-quality seed. Therefore, our results also provide a reference for the selection of mature *Astragalus* seeds.

In this study, population multi-feature modeling and two single features, the R and G values, were used to predict chlorophyll content in *Astragalus* seeds. The multi-feature models were found to be suitable for the prediction of population seed chlorophyll content, market census, government management, and other scenarios. The single feature models were suitable for predicting the chlorophyll content of single *Astragalus* seeds. Using the single-seed R values, we can predict the chlorophyll content of individual seeds, which is of great significance toward improving the quality of seeds in seed batches.

To summarize, digital image analysis technology can be used to determine the relevant features and specific parameters of *Astragalus* seeds. MLR, MLP, and other modeling algorithms combined with the R value and G value single-feature predictions can promote the automation and accuracy of *Astragalus* seed chlorophyll content prediction, which is of great significance to increase the economic benefits of *Astragalus* seeds.

## 5. Patents

The research group has applied for a patent entitled “The construction method and application of a chlorophyll content identification model of *Astragalus* seeds”.

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

**Funding:** This research was funded by the key project “Integration and Industrial Application of Key Technologies in Seed Processing of Genuine Epidemic Medicinal Materials in North China” of the State Administration of Traditional Chinese Medicine “Science and Technology Help Economy 2020” and the City University Cooperation Project “Construction and Key Technology Research of *Astragalus* Breeding Base in Datong City”. The grant numbers were 202004610111024 and 201904710111639.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

**Acknowledgments:** We are very grateful for the equipment support provided by Beijing Youkede Agriculture Co., Ltd.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. Eardley, S.; Bishop, F.L.; Prescott, P.; Cardini, F.; Brinkhaus, B.; Santos-Rey, K.; Vas, J.; von Ammon, K.; Hegyi, G.; Dragan, S.; et al. A systematic literature review of complementary and alternative medicine prevalence in EU. *Forsch. Komplement.* **2012**, *19*, 18–28. [[CrossRef](#)]
2. Yang, J.; Yin, C.; Miao, X.; Meng, X.; Liu, Z.; Hu, L. Rapid discrimination of adulteration in Radix Astragali combining diffuse reflectance mid-infrared Fourier transform spectroscopy with chemometrics. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *248*, 119251. [[CrossRef](#)]
3. Lu, B.; Shi, B.; Xu, L.; Fan, Y.; Lv, B.; Yun, X.; Niu, Q. Effects of Seed Maturation of Sweet Corn on Seed Germination and Seedling Growth. *Seed* **2016**, *35*, 95–99. (In Chinese)
4. Zhan, Y.; Tian, Y.; Ma, J. Study on Relation between Seed Maturity and Germination Rate of Hot Pepper. *J. Jiangxi Agric.* **2011**, *23*, 33–35. (In Chinese)
5. Wilson, H.T.; Khan, O.; Welbaum, G.E. Chlorophyll fluorescence in developing ‘Top Mark’ cantaloupe (*Cucumis melo*) seeds as an indicator of quality. *Seed Technol.* **2014**, *36*, 103–113.
6. Cicero, S.M.; Schoor, R.V.D.; Jalink, H. Use of chlorophyll fluorescence sorting to improve soybean seed quality. *Rev. Bras. Sementes* **2020**, *31*, 145–151. [[CrossRef](#)]
7. Yadav, S.K.; Jalink, H.; Groot, S.P.C.; Van Der Schoor, R.; Yadav, S.; Dadlani, M.; Kodde, J.C. Quality improvement of aged cabbage (*Brassica oleracea* var. capitata) seeds using chlorophyll fluorescence sensor. *Sci. Hortic.* **2015**, *189*, 81–85. [[CrossRef](#)]
8. Deleuran, L.C.; Olesen, M.H.; Boelt, B. Spinach seed quality: Potential for combining seed size grading and chlorophyll fluorescence sorting. *Seed Sci. Res.* **2013**, *23*, 271–278. [[CrossRef](#)]
9. Kenanoglu, B.B.; Demir, I.; Jalink, H. Chlorophyll fluorescence sorting method to improve quality of Capsicum pepper seed lots produced from different maturity fruits. *HortScience* **2013**, *48*, 965–968. [[CrossRef](#)]
10. Jalink, H.; van der Schoor, R.; Birnbaum, Y.E.; Bino, R.J. Chlorophyll fluorescence of *Brassica oleracea* seeds as a non-destructive marker for seed maturity and seed performance. *Seed Sci. Res.* **1998**, *8*, 437–443. [[CrossRef](#)]
11. Groot, S.P.C.; Birnbaum, Y.; Rop, N.; Jalink, H.; Forsberg, G.; Kromphardt, C.; Werner, S.; Koch, E. Effect of seed maturity on sensitivity of seeds towards physical sanitation treatments. *Seed Sci. Technol.* **2006**, *34*, 403–413. [[CrossRef](#)]
12. Yin, J.; Hameed, S.; Xie, L.; Ying, Y. Non-destructive detection of foreign contaminants in toast bread with near infrared spectroscopy and computer vision techniques. *J. Food Meas. Charact.* **2021**, *15*, 189–198. [[CrossRef](#)]
13. Mi, C.; Chen, K.; Zhang, Z. Research on tobacco foreign body detection device based on machine vision. *Trans. Inst. Meas. Control* **2020**, *42*, 2857–2871.
14. Berger, B.; Parent, B.; Tester, M. High-throughput shoot imaging to study drought responses. *J. Exp. Bot.* **2010**, *61*, 3519–3528. [[CrossRef](#)] [[PubMed](#)]
15. Houle, D.; Govindaraju, D.R.; Omholt, S. Phenomics: The next challenge. *Nature Rev. Genet.* **2010**, *11*, 855–866. [[CrossRef](#)] [[PubMed](#)]
16. Montes, J.M.; Melchinger, A.E.; Reif, J.C. Novel throughput phenotyping platforms in plant genetic studies. *Trends Plant Sci.* **2007**, *12*, 433–436. [[CrossRef](#)] [[PubMed](#)]
17. Tanska, M.; Rotkiewicz, D.; Kozirok, W.; Konopka, I. Measurement of the geometrical features and surface color of rapeseeds using digital image analysis. *Food Res. Int.* **2005**, *38*, 741–750. [[CrossRef](#)]
18. Tanska, M.; Ambrosewicz-Walacik, M.; Jankowski, K. Possibility use of digital image analysis for the estimation of the rapeseed maturity stage. *Int. J. Food Prop.* **2017**, *20*, S2379–S2394. [[CrossRef](#)]
19. Ye, F.; Li, L.; Yang, L.; Wang, J.; Sun, Q. Jatropha Seed Selection Technology Based on Machine Vision. *J. China Agric. Univ.* **2016**, *21*, 119–124. (In Chinese)
20. Urena, R.; Rodriguez, F.; Berenguel, M. A machine vision system for seeds germination quality evaluation using fuzzy logic. *Comput. Electron. Agric.* **2001**, *32*, 1–20. [[CrossRef](#)]
21. Granitto, P.M.; Verdes, P.F.; Ceccatto, H.A. Large-scale investigation of weed seed identification by machine vision. *Comput. Electron. Agric.* **2004**, *47*, 15–24. [[CrossRef](#)]
22. Granitto, P.M.; Navone, H.D.; Verdes, P.F.; Ceccatto, H.A. Weed seeds identification by machine vision. *Comput. Electron. Agric.* **2002**, *33*, 91–103. [[CrossRef](#)]
23. Przybyl, K.; Wawrzyniak, J.; Koszela, K.; Adamski, F.; Marzena, G.K. Application of deep and machine learning using image analysis to detect fungal contamination of rapeseed. *Sensors* **2020**, *20*, 7305. [[CrossRef](#)] [[PubMed](#)]
24. Tu, K.; Li, L.; Yang, L.; Wang, J.; Sun, Q. Selection for high quality pepper seeds by machine vision and classifiers. *J. Integr. Agric.* **2018**, *17*, 1999–2006. [[CrossRef](#)]

25. Kara, M.; Sayinci, B.; Elkoca, E.; Ozturk, I.; Ozmen, T.B. Seed size and shape analysis of registered common bean (*Phaseolus vulgaris* L.) cultivars in Turkey using digital photography. *J. Agric. Sci. Tarim Bil. Derg.* **2013**, *19*, 219–234.
26. Huang, K.; Cheng, J. A novel auto-sorting system for Chinese cabbage seeds. *Sensors* **2017**, *17*, 886. [[CrossRef](#)] [[PubMed](#)]
27. Medina, W.; Skurtysb, O.; Aguilera, J.M. Study on image analysis application for identification Quinoa seeds (*Chenopodium quinoa* Willd) geographical provenance. *LWT Food Sci. Technol.* **2009**, *43*, 238–246. [[CrossRef](#)]
28. Tu, K.; Wen, S.; Cheng, Y.; Zhang, T.; Pan, T.; Wang, J.; Wang, J.; Sun, Q. A non-destructive and highly efficient model for detecting the genuineness of maize variety 'JINGKE 968' using machine vision combined with deep learning. *Comput. Electron. Agric.* **2021**, *182*, 106002. [[CrossRef](#)]
29. Krause, W.; Viana, A.P.; Cavalcante, N.R.; Ambrosio, M.; Santos, E.A.; Vieira, H.D. Digital phenotyping for quantification of genetic diversity in inbred guava (*Psidium guajava*) families. *Genet. Mol. Res. GMR* **2017**, *16*, 16019474. [[CrossRef](#)]
30. Pieniazek, F.; Messina, V. Texture and Color Analysis of Freeze-Dried Potato (Cv.Spunta) Using Instrumental and Image Analysis Techniques. *Int. J. Food Prop.* **2016**, *20*, 1422–1431. [[CrossRef](#)]
31. Nambi, V.E.; Thangavel, K.; Shahir, S.; Chandrasekar, V. Color Kinetics during Ripening of Indian Mangoes. *Int. J. Food Prop.* **2016**, *19*, 2147–2155. [[CrossRef](#)]
32. Obulesu, M.; Bhattacharya, S. Color Changes of Tamarind (*Tamarindus Indica* L.) Pulp during Fruit Development, Ripening, and Storage. *Int. J. Food Prop.* **2011**, *14*, 538–549. [[CrossRef](#)]
33. Wang, J. Study on preservation of soybean Germplasm using soybean hard seed. *Soybean Sci.* **1999**, *4*, 351–354. (In Chinese)
34. Cao, B.; Zhai, M.; Guo, J. Study on Vigor Difference of Seeds with Different Hard Degree in *Robinia Pseud. For. Sci.* **2005**, *2*, 42–47. (In Chinese)
35. Tian, J.; Sun, Q.; Wang, J.; Sun, B. Differences of Seed Vigor among Different Levels of Hard Seeds of *Glycyrrhiza uralensis* Fisch. *Plant Physiol. Commun.* **2007**, *2*, 235–240. (In Chinese)
36. He, X. Application of multivariate statistical analysis. *China Stat.* **2015**, *10*, 65. (In Chinese)
37. Kujawa, S.; Nowakowski, K.; Tomczak, R.J.; Dach, J.; Boniecki, P.; Weres, J.; Mueller, W.; Raba, B.; Piechota, T.; Carmona, P.R.C. Neural image analysis for maturity classification of sewage sludge composted with maize straw. *Comput. Electron. Agric.* **2014**, *109*, 302–310. [[CrossRef](#)]
38. Boniecki, P.; Koszela, H.; Piekarska, B.J.; Weres, M.; Zaborowicz, S.; Kujawa, A.; Majewski, B.R. Neural identification of selected apple pests. *Comput. Electron. Agric.* **2015**, *110*, 9–16. [[CrossRef](#)]
39. Pan, W.; Ma, W.; Zhen, J.; Geng, S. Chlorophyll fluorescence of tobacco seeds as marker of seed maturity. *Southwest China J. Agric. Sci.* **2016**, *29*, 966–969. (In Chinese)