Prediction of Buckwheat Maturity in UAV-RGB Images Based on Recursive Feature Elimination Cross-Validation: A Case Study in Jinzhong, Northern China

Jinlong Wu 1,2, Decong Zheng 1, Zhiming Wu 1, Haiyan Song 1,* and Xiaoxiang Zhang 1

1 College of Agricultural Engineering, Shanxi Agricultural University, Jinzhong 030801, China
2 College of Information Science and Engineering, Shanxi Agricultural University, Jinzhong 030801, China
* Correspondence: yybbao@sxau.edu.cn

Abstract: Buckwheat is an important minor grain crop with medicinal and edible functions. The accurate judgment of buckwheat maturity is beneficial to reduce harvest losses and improve yield. With the rapid development of unmanned aerial vehicle (UAV) technology, it has been widely used to predict the maturity of agricultural products. This paper proposed a method using recursive feature elimination cross-validation (RFECV) combined with multiple regression models to predict the maturity of buckwheat in UAV-RGB images. The images were captured in the buckwheat experimental field of Shanxi Agricultural University in Jinzhong, Northern China, from September to October in 2021. The variety was sweet buckwheat of “Jinqiao No. 1”. In order to deeply mine the feature vectors that highly correlated with the prediction of buckwheat maturity, 22 dimensional features with 5 vegetation indexes, 9 color features, and 8 texture features of buckwheat were selected initially. The RFECV method was adopted to obtain the optimal feature vector dimensions and combinations with six regression models of decision tree regression, linear regression, random forest regression, AdaBoost regression, gradient lifting regression, and extreme random tree regression. The coefficient of determination ($R^2$) and root mean square error (RMSE) were used to analyze the different combinations of the six regression models with different feature spaces. The experimental results show that the single vegetation index performed poorly in the prediction of buckwheat maturity; the prediction result of feature space “5” combined with the gradient lifting regression model performed the best; and the $R^2$ and RMSE were 0.981 and 1.70 respectively. The research results can provide an important theoretical basis for the prediction of the regional maturity of crops.

Keywords: recursive feature elimination cross-validation; UAV-RGB images; prediction of buckwheat maturity; feature vectors; regression models

1. Introduction

Buckwheat is an important minor grain crop in China. It is helpful to lower blood pressure, control diabetes, and improve digestion and cholesterol level [1]. The harvest time of the crop has a significant impact on its yield and quality; early or late harvest is not conducive to the high yield and income of the crop [2]. The biggest characteristic of buckwheat during its growth is the long overlapping period, in which the phenotypic characteristics of buckwheat change greatly. The overall growth period can be divided into four shorter sub-periods: budding, flowering, growth, and maturity. Therefore, the prediction of buckwheat maturity can be realized according to the phenotypic characteristics in these different periods. The accurate judgment of buckwheat maturity is beneficial to the advance scheduling of harvesters, and can effectively reduce harvest losses and improve yield. The traditional methods for evaluating buckwheat maturity are mainly based on field measurement and the experience of farmers. In large-scale planting, subjective misjudgment, manpower, and material resources waste are all factors. In recent years, the development of UAV has provided a new idea for the prediction of crop maturity, and the...
UAV technology can be used to further explore the features highly related to buckwheat planting information. In the flowering period, grains with different maturity levels will appear on the same buckwheat plant, and plants with different maturity levels will also appear on the same plot. If the harvest time is too early, most of the grains will be immature, and if the harvest time is too late, the grains will fall off, which will cause great economic losses. Therefore, accurate calibration of buckwheat maturity is of great significance to ensure buckwheat yield. Generally, when the grain maturity of a single buckwheat plant reaches 75–80%, the color of the grain turns brown or gray [3], which means that the single buckwheat plant is mature. Since UAV photography can only obtain the canopy image of buckwheat, it cannot accurately describe the mature state of each buckwheat plant, and then the overall maturity of each buckwheat plot can be calibrated. The empirically based maturation calibration process allows an error of one to two days.

UAV technology has been widely used in crop growth monitoring [4,5], pest and disease control [6], soil analysis and planning [7,8], precision fertilization [9], and other aspects, and the prediction of crop maturity is a major application of crop growth monitoring. At present, there have been studies on UAV remote sensing platforms to predict the maturity of agricultural products. Rodrigo Trevisan et al. [10] used convolutional neural networks (CNN) to predict the maturity of soybean in airborne RGB images, and the prediction results of $\text{RMSE}$ can reach 2.0 days. Jing Zhou et al. [11] predicted the maturity of soybean in airborne multispectral images by partial least square regression (PLSR), and the prediction results of $R^2$ and $\text{RMSE}$ were 0.81 and 1.4 days, respectively. Neil Yu et al. [12] developed a dual-camera, high-throughput phenotype platform mounted on a UAV and used the random forest method to measure soybean maturity; the prediction accuracy was 93%. However, most of the above research is aimed at the prediction of maturity in UAV images such as soybeans and wheat, and research on maturity prediction of minor grain crops such as buckwheat are fewer. Generally, methods of judging buckwheat maturity are mainly based on its phenotypic characteristics. At present, research on crop phenotype in UAV images primarily focus on indexes of leaf area index (LAI), leaf dry matter (LDM) [13], plant density (PD) [14], yield prediction [15], and above ground biomass (AGB) [16]. At the same time, most of the crop phenotype research based on UAV consists of multispectral and hyperspectral images. Experimental results show that the inversion of physical and chemical parameters can be realized by using the strong correlation between the fixed bands of multispectral and hyperspectral spectra and the biochemical moisture and pigment of plants. Since buckwheat is widely planted in China, and the UAVs equipped with multispectral and hyperspectral cameras are expensive, the data processing is more complicated, making it difficult to apply them to practical production. UAVs equipped with an RGB camera have been widely used in the field of crop classification and recognition due to the camera’s low cost and easy access. Hence, this study deeply explores the potential value of UAV-RGB images of buckwheat for maturity prediction by extracting multiple feature vectors from the images, obtaining the optimal combination of feature vectors and the optimal regression model, and demonstrating the prediction of buckwheat maturity in UAV-RGB images.

In this study, a method for buckwheat maturity prediction was proposed based on easily accessible and low-cost high-resolution UAV-RGB images. The main contributions are as follows:

1. The UAV-RGB images of buckwheat are collected periodically in the overlapping period, and then the vegetation indexes, color and texture features are extracted.
2. In view of interference information, such as bare ground in the images, and the subjectivity of feature selection in current research, correlation analysis and recursive feature elimination cross-validation methods are adopted for feature selection. Combining the selected features with multiple regression models allows the optimal combination of feature vectors and the optimal regression prediction model to be determined.
3. We evaluate the accuracy of the prediction model proposed in this paper in order to provide reference for UAV remote sensing detection for crops.
2. Materials and Methods

2.1. Research Area

The experiment was carried out in 2021 in the buckwheat research experimental field (37°26′2.4″ N–37°26′6″ N, 112°35′34.8″–112°35′42.0″ E) of Shanxi Agricultural University in Shenfeng village, Taigu District, Jinchong City, Shanxi Province. The sowing time of buckwheat was July 2, and the variety was sweet buckwheat “Jinqiao No. 1”. During the experiment, the highest temperature was 31 degrees, the lowest temperature was 6 degrees, and the average rainfall was 119.25 mm. The total area of the experimental field was 13,021 m², which was divided into 35 experimental plots with an average area of 3 × 3 m for each plot (Figure 1).

![Figure 1](image-url)

**Figure 1.** Geographical location of the research area and the buckwheat experimental plots. (a) Geographical location of the research area. (b) Buckwheat experimental plots.

2.2. UAV Images Acquisition

The aerial image acquisition equipment is Dajiang PHANTOM 4RTK UAV (equipped with 20-megapixel 1-inch CMOS sensor), which is carried with the RTK module, compatible with high-precision GNSS mobile stations, and has centimeter level positioning capability. Setting of aerial photography parameters: the flight height is 30 m, the flight speed is 2.6 m/s, the GSD is 0.82 cm/pixel, the heading overlap rate is 80%, and the side overlap rate is 70%. The collection time is mainly after the flowering period of buckwheat, and the specific collection dates are 9, 15, 21, 28 September and 8, 12 October in 2021, respectively. Figure 2 shows the images collected at the same position of the buckwheat experimental field in different periods. Figure 2a,b show the squaring period of buckwheat, during which the crop sizes are small and the crop rows are clearly visible. Figure 2c,d show the flowering period of buckwheat, during which buckwheat grows rapidly, the canopy is mainly the color of flowers, and the crop rows are gradually less obvious. Figure 2e,f show the growth period and maturity period of buckwheat, respectively, and it is clearly evident that there are mature grains on the crops, and the overall color of the canopy has started to darken and turn brown. The results show that buckwheat plots at the same location in different periods have obvious differences in canopy color and texture. Therefore, the initial feature vector can be constructed from the color and texture features of buckwheat plots to predict the maturity of buckwheat.
2.3. Image Segmentation

In the growth of buckwheat, it is difficult to achieve the segmentation of single buckwheat because of the high planting density and the serious overlapping of leaves. Therefore, image segmentation can be used to separate the canopy area from the background area. Since there are great color differences between the crop area and the background, the canopy area can be segmented by color features. After analysis, the image segmentation results of the original R, G, and B channels are not ideal, so the commonly used vegetation indexes [17] (VI) in UAV remote sensing are introduced for image segmentation. Through experimental comparison, the excess green index (ExG) has better segmentation performance; its formula is shown in (1):

\[
ExG = 2G - R - B
\]  

(1)

In order to realize the segmentation of the buckwheat canopy, the original RGB image was converted into an ExG gray-scale image, and then the K-means algorithm was used to cluster the image pixels into buckwheat canopy and background (“0” represents the background, “1” represents the canopy). Then, morphological open and close operations [18] were applied for noise removal and holes filling, and the final binary image was the segmentation result of buckwheat canopy.

2.4. Feature Extraction

This study constructed prediction features from three aspects: vegetation indexes, color features, and texture features. As the research object in this paper is the UAV-RGB image, five vegetation indexes based on the R, G, and B bands were selected, including the normalized green red difference index (NGRDI) [19], Green Leaf Algorithm (GLA), Visible Atmospherically Resistant Index (VARI), ExG, and Normalized Difference Yellow Index (NDYI), and the corresponding formulas are shown in Table 1.
Table 1. Vegetation indices.

| Vegetation Indexes Name                          | Abbreviation | Formula                                      |
|-------------------------------------------------|--------------|----------------------------------------------|
| Normalized green red difference index            | NGRDI        | \((g - r)/(g + r)\)                        |
| Green Leaf Algorithm                             | GLA          | \((2g - r - b)/(2g + r + b)\)               |
| Visible Atmospherically Resistant Index          | VARI         | \((g - r)/(g + r - b)\)                    |
| Excess green index                               | ExG          | \((2g - r - b)/(g - b)/(g + b)\)          |
| Normalized Difference Yellow Index               | NDYI         |                                            |

Note: R, G, and B, respectively, represent the reflection values of the R, G, and B channels of the UAV images.

The UAV-RGB images acquired at the same time and location have great differences due to the influence of illumination. Therefore, the RGB images can be converted into multiple color spaces to reduce the influence of illumination on the prediction results. By analyzing the common color space HSV [20], HLS [21], and Lab [22], color feature vectors that have strong correlation with the prediction results of buckwheat maturity were explored. In the test, HSV H, HSV S, and HSV V represent the H, S, and V components of the HSV color space, respectively; Lab L, Lab a, and Lab b represent the L, a, and b components of the Lab color space, respectively; and HLS H, HLS L, and HLS S represent the H, L, and S components of the HLS color space, respectively. Thus, nine color features of HSV, HLS, and Lab color spaces were selected for the alternative feature space used for the prediction of buckwheat maturity.

As the buckwheat rows in the squaring period, the shape of the flowers in the flowering period, the number of grains in the growth period, and the maturity period of the UAV-RGB image are quite different in texture, texture features can effectively predict the maturity of buckwheat. In this paper, eight common texture features of each buckwheat plot were selected for feature space construction. Including six gray-level co-occurrence matrix values [23], with local binary pattern (LBP) [24] and Gabor [25] texture feature mean among them, the gray-level co-occurrence matrix values are Homogeneity (HOM), Contrast (CON), Dissimilarity (DIS), Entropy (ENT), Angular Second Moment (ASM), and Correlation (COR).

2.5. Maturing Period Calibration and Correction

2.5.1. Maturing Period Calibration

The calibration of the buckwheat maturity period can be realized according to the acquisition time of UAV-RGB images in each period. That is, the number of days between the harvest date and the current image capture date of buckwheat is defined as the maturity period, which can be denoted as \(MD_i\), and it facilitates the establishment of the later regression model. The harvest date of buckwheat in this experiment was 15 October 2021, and the capture date of buckwheat images in the different periods was 9, 15, 21, 28 September and 8, 12 October 2021, respectively; therefore, the corresponding interval time of 36 days, 30 days, 24 days, 17 days, 7 days, and 3 days can be preliminarily considered as the buckwheat maturity period, and its physical meaning is the number of days in the maturity period.

2.5.2. Maturing Period Correction

Most buckwheat has matured in the harvest time, but there are still a few immature grains in the calibration process, which will reduce the accuracy of the later model training. Figure 3 shows the images taken at different locations on the same date of buckwheat maturity. It can be seen in Figure 3a that mature grains account for a relatively small proportion at the end of the flowering period; In Figure 3b, most of the grains are mature, and the overall color is brown. Therefore, it is necessary to recalibrate the maturity period according to the color characteristics of the actual canopy in each buckwheat plot. In order to reduce the difficulty of calibration, it is generally considered that the error before and after the actual maturity of buckwheat in the harvest time should not exceed three days. In the process of calibration, the proportion of brown pixels can be approximately regarded as...
the maturity index, which can be marked as $D_i$, and the value range is $(-1.5, 1.5)$; hence, the revised maturity date is $\text{MMD}_i$ as shown in Formula (2).

$$\text{MMD}_i = \text{MD}_i + D_i$$  \hspace{1cm} (2)

Figure 3. Images taken at different locations on the same date during the buckwheat maturity period. (a) Partial immature. (b) Basically mature.

2.6. Feature Selection

In order to explore features that can effectively represent the maturity information of buckwheat, this paper extracts vegetation indexes, texture features, and color features of the UAV-RGB image to initialize the feature space. Feature selection can obtain the best combination of feature vectors to make the prediction results the best and ensure the prediction accuracy of the model while reducing the amount of calculation and the difficulty of model learning. The Pearson correlation coefficient can effectively measure the linear correlation among feature vectors and effectively reduce the redundancy of the feature space by removing feature vectors with a high correlation coefficient. Therefore, this paper uses the Pearson correlation coefficient for feature space analysis.

When performing regression analysis on the optimized feature space, there is a relationship of rising first and falling later between the dimension of the feature vectors and the prediction result of the model; thus, the dimension of the feature vectors will seriously affect the prediction accuracy of the model. Therefore, it is necessary to search the optimal dimensions and combinations of feature vectors that make the prediction results the best. Encapsulated feature selection integrates a regression model into the process of feature selection, takes the cross-validation results as evaluation criteria, and selects feature vectors with high contribution values as the final results. Common encapsulated feature selection methods include stability selection, sequential feature selection, and recursive feature elimination [26]. In this paper, recursive feature elimination (RFE) was used to select the optimal combinations of feature vectors, and its working principle is to select features by continuously reducing the volume of the feature sets through recursive methods. The implementation steps of the algorithm are as follows:

1. The initial feature space was the combination of all feature vectors, through which the regression model was trained. The importance of each feature was determined by the attributes of the correlation coefficient and feature importance;
2. The features with the lowest importance were removed from the current feature combinations, and then the process of feature pruning was repeated recursively until the set number of features was reached;
3. Since the recursive feature elimination method needs to manually set the number of features, it is unable to automatically determine the optimal number of features. Therefore, cross-validation can be introduced into the recursive feature elimination, that is, recursive feature elimination with cross-validation (RFECV) [27]. It can automatically determine the optimal number and combinations of features by using the regression model in the process of cross-validation, making the prediction results of the model optimal. In the process of FRFECV, this paper used five-fold cross-validation to select the number and combinations of features.
2.7. Regression Model Establishment

The processes of feature extraction, model establishment, and data analysis of the UAV-RGB buckwheat images was realized in Python 3.8.8. The computer is configured as Inter (R) Core (TM) i7-6700@3.4 GHz, and the memory is 8 G. In this paper, decision tree regression, linear regression, random forest regression, AdaBoost regression, gradient boosting regression, and extra tree regression [28] were used for comparative experiments to verify the application of UAV-RGB images in predicting buckwheat maturity. The above regression models were integrated into the process of FRFECV to obtain the optimal dimensions and combinations of feature vectors.

2.8. Model Evaluation Indexes

In order to effectively evaluate the importance of each feature vector in regression analysis, the permutation feature importance (PFI) index was introduced. It indicates the decrease of score in the regression model when the value of a single feature vector is randomly disturbed. The PFI score represents the dependence degree of the model on a feature vector, and the prediction performance of different features in model training can be determined by the sorting result of PFI. The Coefficients of determination ($R^2$) and Root Mean Square Error (RMSE) [29] were used to verify the prediction accuracy of the model. The formulas are as follows (3)–(5):

$$PFI_j = s - \frac{1}{K} \sum_{k=1}^{K} s_{k,j}$$

(3)

where $PFI_j$ represents the importance of the arrangement characteristics corresponding to the $j$th feature vector; $S$ is the reference value of a specific regression model after feature vector learning, and the reference value of the regression model is calculated by $R^2$ of the model; $K$ indicates the number of times that the feature vector value is scrambled.

$$R^2 = \frac{\sum_{i=0}^{n} (X_i - \bar{X})^2 (Y_i - \bar{Y})^2}{n\sum_{i=0}^{n} (X_i - \bar{X})^2 \sum_{i=0}^{n} (Y_i - \bar{Y})^2}$$

(4)

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=0}^{n} (Y_i - X_i)^2}$$

(5)

where $X_i$ and $Y_i$, respectively, represent the estimated and measured values of the days from maturity; $\bar{X}$ and $\bar{Y}$ represent the mean value of $X_i$ and $Y_i$; $n$ is the number of samples.

3. Results

3.1. Feature Selection Results

The Pearson correlation coefficient method was used to analyze the 22 extracted feature vectors. Figure 4 is the thermodynamic diagram of the correlation coefficient feature vectors, and the darker the color is, the higher the correlation is between feature vectors. It can be seen from Figure 4 that the correlation coefficient between HSV_H and HLS_H is 1, as the calculation formulas of component Hue in HSV and HLS are the same, then feature HSV_H can be removed. The correlation coefficients between GLA and ExG, HSV_S, Lab_b are 0.996, 0.933 and 0.941, respectively, and their correlation coefficients all exceed 0.9, as their contribution in regression analysis are similar; thus, Lab_b, ExG and HSV_S can be removed and GLA retained. Similarly, 7 feature vectors were removed finally, and the original 22 feature vectors were reduced to 15: vegetation indexes (NGRDI, GLA, VARI), color features (Lab_a, HLS_H, HLS_L, HLS_S, HSV_S, HSV_V), texture features (LBP, Gabor, HOM, CON, COR, ASM).

RFECV was used to twice optimize the feature space. In the process of cross-validation, six regression models of decision tree regression, linear regression et al. were embedded in RFE, and the relationship between the regression prediction values and feature vectors
was used for optimal selection. Figure 5 shows the optimal number of feature vectors corresponding to different regression models, in which the abscissa is the dimension of feature vectors, and the ordinate is the cross-validation accuracy of the prediction models. It can be seen from Figure 5 that when the number of the selected features exceeds a certain value, the prediction results of all models tend to be stable, and the increase of the number of feature vectors will not improve the accuracy of the models, but will greatly increase the amount of computation. Therefore, it is unreasonable to blindly increase the number of feature vectors.

In the decision tree regression model, when the number of feature vectors reaches 2, the prediction result is optimal, and the corresponding feature vectors are color feature (HLS_S) and texture feature (ASM), and the feature space is defined as “1”. In the linear regression model, the prediction result is optimal when the number of feature vectors reaches 4, and the corresponding feature vectors are vegetation indexes (NGRDI, VARI) and texture features (ASM, COR), and it is defined as feature space “2”. In the random forest regression model, when the number of feature vectors reaches 3 after feature selection by RFECV, the result is optimal, and the corresponding feature vectors are texture features (ASM, COR) and color feature (HLS_S), and it is defined as feature space “3”. In the AdaBoost regression model, the optimal feature vector dimension is 4, which are texture features (CON, ASM and COR) and color feature (HLS_S), and it is defined as feature space “4”. In the gradient lifting regression model, when the number of feature vectors reaches 5, it reaches the optimum, which is defined as feature space “5”, and the feature space includes vegetation index (VARI), texture features (CON, ASM and COR) and color feature

Figure 4. Thermodynamic diagram of the correlation coefficient feature vectors.
(HLS_S). The extreme random tree regression model needs 10 feature vectors to reach the optimum, which is defined as feature space “6”, and it includes vegetation index (NGRDI), texture features (CON, ASM, COR, HOMO, Gabor) and color features (HLS_L, HLS_H, HLS_S, Lab_a).

Figure 5. The optimal number of corresponding feature vectors for different regression models.

After feature selection by Pearson correlation coefficient and RFECV, the optimal dimensions and combinations of feature vectors that correspond to different regression models were obtained finally. Figure 6 shows the PFI values of feature vectors corresponding to different regression models, in which the abscissa is the importance parameter of the arrangement characteristics, and the ordinate is the feature vectors. The boxplot can reflect the distribution of different feature vectors on PFI values. It can be seen from Figure 6 that the importance of feature vector PFIs are different in the six regression models. The importance of the arrangement characteristics of texture feature COR in decision tree regression, random forest regression, AdaBoost regression, and gradient lifting regression are high. The texture features of LBP and Gabor are difficult to reflect the texture characteristics of the whole image on a single value, and the importance is much lower in all six regression models. The experimental results show that the importance of the permutation characteristics of each feature vector is basically consistent with the feature space selected by RFECV, which verifies the effectiveness of our method.

Figure 6. Cont.
Figure 6. Feature vector PFI values of different regression models. (a) Decision tree regression. (b) Linear regression. (c) Random forest regression. (d) AdaBoost regression. (e) Gradient lifting regression. (f) Extreme random tree regression.

3.2. Prediction Algorithm Selection Results

After feature selection, feature vector optimization, and combination, the optimal feature vector combinations for each model were obtained. In order to evaluate the matching results between the regression models and the feature spaces, vegetation indexes, texture features, color features, and all features, respectively, combined with different regression models were conducted for contrast experiments. The regression analysis models with the prediction results of the corresponding feature space are shown in Table 2, and the results are the average of several experiments.

According to Table 2, although the prediction results of the six regression models are good for all feature spaces, they are not optimal compared with other feature spaces, which indicates that more feature vectors does not mean better. It is necessary to find the optimal feature space for different regression models. The prediction results of the vegetation index in the six regression models are all poor; hence, it can be considered that the vegetation index alone cannot accurately predict buckwheat maturity. Compared with the vegetation index, the regression results of different regression models are better for texture features and color features. In the decision tree regression model, the feature space with the highest $R^2$ is “1”, and its $RMSE$ is 2.27; the regression prediction result is the best compared with other model spaces, which is the same as the optimization result of RFECV. In the linear regression model, its optimal space is feature space “2”. There is a slight decrease in $R^2$ and $RMSE$ compared with all feature spaces, but it can greatly reduce the calculation amount of regression analysis. To summarize, the prediction results of random forest regression in
feature space “3”, “4”, and “5” are the same, and $R^2$ and RMSE are the highest compared with other feature spaces. Gradient lifting regression performs best in the same feature space, and the result of feature space “5” is the best compared with other feature spaces. Therefore, the method combining feature space “5” with gradient lifting regression can be used to realize the prediction of buckwheat maturity based on UAV-RGB images. The optimal determination coefficient ($R^2$) and root mean square error (RMSE) are 0.981 and 1.70, respectively, and the results basically meet the results of RFECV optimization.

Table 2. Regression analysis model and corresponding feature space prediction results.

| Feature Spaces | Decision Tree Regression | Linear Regression | Random Forest Regression | AdaBoost Regression | Gradient Lifting Regression | Extreme Random Tree Regression |
|----------------|--------------------------|-------------------|--------------------------|----------------------|-----------------------------|-------------------------------|
|                | $R^2$ | RMSE | $R^2$ | RMSE | $R^2$ | RMSE | $R^2$ | RMSE | $R^2$ | RMSE | $R^2$ | RMSE |
| Vegetation Indexes | 0.354 | 9.56 | 0.624 | 7.30 | 0.612 | 7.42 | 0.683 | 6.69 | 0.687 | 6.66 | 0.616 | 7.34 |
| Texture Features | 0.869 | 4.31 | 0.856 | 4.52 | 0.868 | 4.33 | 0.907 | 3.63 | 0.892 | 3.91 | 0.842 | 4.71 |
| Color Features | 0.922 | 3.27 | 0.851 | 4.59 | 0.848 | 4.64 | 0.928 | 3.17 | 0.937 | 2.98 | 0.792 | 5.33 |
| Feature Space 1 | 0.964 | 2.27 | 0.745 | 6.01 | 0.949 | 2.68 | 0.955 | 2.52 | 0.962 | 2.33 | 0.929 | 3.16 |
| Feature Space 2 | 0.882 | 4.08 | 0.902 | 3.73 | 0.868 | 4.32 | 0.919 | 3.38 | 0.921 | 3.35 | 0.867 | 4.27 |
| Feature Space 3 | 0.951 | 2.64 | 0.855 | 4.54 | 0.949 | 2.68 | 0.960 | 2.39 | 0.971 | 2.02 | 0.924 | 3.00 |
| Feature Space 4 | 0.960 | 2.35 | 0.856 | 4.53 | 0.949 | 2.68 | **0.965** | **2.21** | 0.977 | 1.80 | 0.927 | 2.92 |
| Feature Space 5 | 0.949 | 2.68 | 0.855 | 4.54 | 0.949 | 2.68 | 0.959 | 2.40 | **0.981** | **1.70** | 0.926 | 3.21 |
| Feature Space 6 | 0.959 | 2.39 | 0.887 | 3.99 | 0.948 | 2.71 | 0.957 | 2.47 | 0.975 | 1.90 | 0.925 | 3.21 |
| All Feature Space | 0.954 | 2.54 | 0.934 | 3.06 | 0.948 | 2.72 | **0.966** | **2.20** | 0.980 | 1.70 | **0.928** | **3.08** |

4. Discussion

In this paper, the Pearson correlation coefficient was used for preliminary dimensionality reduction of the feature space. We reduced the original 22 feature vectors to 15, proving the effectiveness of the Pearson correlation coefficient, which included vegetation indexes (NGRDI, GLA, VARI), color features (Lab_a, HLS_H, HLS_L, HLS_S, HSV_S, HSV_V), and texture features (LBP, Gabor, HOM, CON, COR, ASM).

Regression models were integrated into the process of feature selection, in which the cross-validation results were used as the evaluation criteria, and the feature vectors with higher contribution values were selected as the results. Through RFECV, the optimal number and combinations of feature vectors corresponding to different regression models were obtained. Moreover, six feature spaces were defined. The experimental results show that the method of combining feature space “5” (vegetation index VARI; texture features CON, ASM and COR; and color feature HLS_S) with gradient lifting regression can realize the prediction of buckwheat maturity, and its $R^2$ and RMSE are 0.981 and 1.70, respectively, when achieving the best result in all the combinations.

Most of the images used for crop maturity predictions are satellite remote sensing images [30,31] (research on rice and corn) and UAV images. As there has been little previous work applied to crop maturity predictions in UAV images, let alone the applications in buckwheat, this work can be compared to the prediction results for soybean. Two complementary convolutional neural networks (CNN) were developed to predict the maturity date in reference [10], and the data were acquired from three growing seasons in the USA and Brazil, and the RMSE was 2.0 days. In reference [11], the prediction of soybean maturity dates was realized using UAV multispectral imagery, the experiment was conducted at an experimental field at the University of Missouri, Novelty, Missouri, United States, and the $R^2$ and RMSE were 0.81 and 1.4 days, respectively. Compared with state-of-arts on soybean, in this paper, the improvement of $R^2$ is obvious and the value of RMSE can meet the general error requirement. Although the use of deep learning can perform well in predictions, it must depend on a large amount of data to guarantee the robustness of the model. The regression analysis model used in this paper can also meet the prediction requirements with small data sets. Furthermore, the optimal feature spaces
corresponding to different regression analysis methods can be obtained automatically. Moreover, when conducting feature dimension reduction with PCA (Principal Component Analysis), the features obtained usually have no practical significance, while the optimal feature space obtained in our method has practical physical significance that can represent the characteristics of buckwheat maturity. In addition, it can also reduce the amount of calculation for feature extraction in later detection.

Analyzing the optimal feature space showed the optimal feature space with different regression models contains texture features (ASM and COR) and color features (HLS_S), and the vegetation index performs poorly. Therefore, the correlation between texture features, color features, and maturity can be further explored. Remote sensing technology can only obtain the phenotypic information of crops, and it is difficult to describe the internal mechanism changes of crops. In reference [32], visible and near-infrared hyperspectral imaging was employed to evaluate the maturity stage and moisture content of fresh okra fruit precisely, and the physicochemical analysis indicated that there was a negative correlation between maturity and moisture content. To assess the viability of growing cover crops in Denmark [33], a phenology model was developed and applied to predict the harvest date of spring barley and winter wheat. Therefore, the phenological information, crop growth information, and other chemical indicators can be introduced to the prediction model of crop maturity in the future.

5. Conclusions

In this study, RFECV was used to integrate six regression models of decision tree regression and random forest regression et al. into feature selection, and the optimal dimension of the feature vector was obtained during the process of cross-validation, and the feature vectors that can effectively predict the maturity of buckwheat in UAV-RGB images were mined in depth. The traditional prediction of buckwheat maturity usually depends on manual judgment to determine the corresponding harvest (sectional harvest and one-time harvest). However, when the planting area is larger, it is difficult to evaluate the maturity of the field manually. Then it is valuable to use UAV to predict the maturity of buckwheat. Moreover, the cost of acquiring and processing images based on RGB and UAV in this paper is relatively low, which can be widely promoted.

For off-line prediction of crops maturity based on UAV images, priority should be given to improving the performance of the regression models, while calculation requirements need not be considered to some extent. However, for on-line prediction, priority should be given to calculation requirements, and the stability of the system needs to be guaranteed. The premise of reducing the amount of computation is to ensure that the performance of the model will not be greatly affected, and finally to provide a theoretical basis for the application of UAV in agricultural on-line detection. Furthermore, the time series images were collected at a fixed flight altitude of the same field. The impact of different flight altitudes and imaging resolutions on the experimental results were not considered. The follow-up work can continue to carry out in-depth mining of the UAV-RGB images, taking into account the spatial location information and weather information. Moreover, it is valuable and necessary for further research on buckwheat phenotype analysis, such as chlorophyll content, nitrogen content, and yield estimation.

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

Funding: This research was funded by National Key Research and Development Program, grant number 2021YFD1600602-09; China Modern Agricultural Industrial Technology System, grant number CARS-07-D-2; State Key Laboratory of Sustainable Dry Land Agriculture (in preparation), Shanxi Agricultural University, grant number 202003-7.
Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank the anonymous reviewers for their constructive comments, which considerably improved the quality of the final manuscript.

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

References

1. Ninomiya, K.; Yamaguchi, Y.; Shinmachi, F.; Kumagai, H.; Kumagai, H. Suppression of postprandial blood glucose elevation by buckwheat (Fagopyrum esculentum) albumin hydrolysate and identification of the peptide responsible to the function. Food Sci. Hum. Wellness 2022, 11, 992–998. [CrossRef]

2. Zhang, J.J.; Wei, Q.Q.; Xiong, S.P.; Shi, L.; Ma, X.M.; Du, P.; Guo, J.B. A spectral parameter for the estimation of soil total nitrogen and nitrate nitrogen of winter wheat growth period. Soil Use Manag. 2021, 37, 698–711. [CrossRef]

3. Hennessy, P.J.; Esau, T.J.; Schumann, A.W.; Zaman, Q.U.; Corssadden, K.W.; Faroque, A.A. Evaluation of cameras and image distance for CNN-based weed detection in wild blueberry. Smart Agric. Technol. 2022, 2, 100030. [CrossRef]

4. Zhu, W.X.; Rezaei, E.E.; Nouri, H.; Sun, Z.G.; Li, J.; Yu, D.Y.; Siebert, S. UAV-based indicators of crop growth are robust for distinct water and nutrient management but vary between crop development phases. Field Crops Res. 2022, 284, 108582. [CrossRef]

5. Huang, H.S.; Yang, A.Q.; Zhuang, J.J.; Hou, C.J.; Tan, Z.P.; Dananjayan, S.; He, Y.; Guo, Q.W.; Luo, S.M. Deep color calibration for UAV imagery in crop monitoring using semantic style transfer with local to global attention. Int. J. Appl. Earth Obs. Geoinf. 2021, 104, 102590. [CrossRef]

6. Tetila, E.C.; Machado, B.B.; Astolfi, G.; Belete, N.A.D.S.; Amorim, W.P.; Roel, A.R.; Pistori, H. Detection and classification of soybean pests using deep learning with UAV images. Comput. Electron. Agric. 2020, 179, 105836. [CrossRef]

7. Bertalan, L.; Holb, I.; Patakai, A.; Négyesi, G.; Szabó, G.; Szaloki, A.K.; Szabó, S. UAV-based multispectral and thermal cameras to predict soil water content—A machine learning approach. Comput. Electron. Agric. 2022, 200, 107262. [CrossRef]

8. Das, S.; Christopher, J.; Choudhury, M.R.; Apan, A.; Chapman, S.; Menzies, N.W.; Dang, Y.P. Evaluation of drought tolerance of wheat genotypes in rain-fed sodic soil environments using high-resolution UAV remote sensing techniques. Biosyst. Eng. 2022, 217, 68–82. [CrossRef]

9. Song, C.C.; Zhou, Z.Y.; Zhang, Y.; Zhao, L.L.; Sun, W.W.; Luo, X.W.; Jiang, R.; Ming, R.; Zang, Y.; Zi, L.; et al. Variable-rate control system for UAV-based granular fertilizer spreader. Comput. Electron. Agric. 2021, 180, 105832. [CrossRef]

10. Trevisan, R.; Pérez, O.; Schmitz, N.; Diers, B.; Martin, N. High-Throughput Phenotyping of Soybean Maturity Using Time Series UAV Imagery and Convolutional Neural Networks. Remote Sens. 2020, 12, 3617. [CrossRef]

11. Zhou, J.; Yunghluth, D.; Yong, C.N.; Scaboo, A.; Zhou, J.F. Estimation of the Maturity Date of Soybean Breeding Lines Using UAV-Based Multispectral Imagery. Remote Sens. 2019, 11, 2075. [CrossRef]

12. Yu, N.; Li, L.J.; Schmitz, N.; Tian, L.F.; Greenberg, J.A.; Diers, B.W. Development of methods to improve soybean yield estimation and predict plant maturity with an unmanned aerial vehicle based platform. Remote Sens. Environ. 2016, 187, 91–101. [CrossRef]

13. Zhang, J.Y.; Qiu, X.L.; Wu, Y.T.; Zhu, Y.; Cao, Q.; Liu, X.J.; Cao, W.X. Combining texture, color, and vegetation indices from fixed-wing UAS imagery to estimate wheat growth parameters using multivariate regression methods. Comput. Electron. Agric. 2021, 185, 106138. [CrossRef]

14. Randelovic, P.; Dordevic, V.; Milic, S.; Balesevic-Tubic, S.; Petrovic, K.; Miladinovic, J.; Dukic, V. Prediction of Soybean Plant Density Using a Machine Learning Model and Vegetation Indices Extracted from RGB Images Taken with a UAV. Agronomy 2020, 10, 1108. [CrossRef]

15. Shafiee, S.; Lied, L.M.; Burud, I.; Dieseth, J.A.; Alsheikh, M.; Lillemo, M. Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. Comput. Electron. Agric. 2021, 183, 106036. [CrossRef]

16. Grüner, E.; Wachendorf, M.; Astor, T. The potential of UAV-borne spectral and textural information for predicting aboveground biomass and N fixation in legume-grass mixtures. PLoS ONE 2020, 15, e0234703. [CrossRef]

17. Wang, N.; Guo, Y.C.; Wei, X.; Zhou, M.T.; Wang, H.J.; Bai, Y.B. UAV-based remote sensing using visible and multispectral indices for the estimation of vegetation cover in an oasis of a desert. Ecol. Indic. 2022, 141, 109155. [CrossRef]

18. Qiao, L.; Gao, D.H.; Zhao, R.M.; Tang, W.J.; An, L.L.; Li, M.Z.; Sun, H. Improving estimation of LAI dynamic by fusion of morphological and vegetation indices based on UAV imagery. Comput. Electron. Agric. 2022, 192, 106603. [CrossRef]

19. Lu, J.S.; Cheng, D.L.; Geng, C.M.; Zhang, Z.T.; Xiang, Y.Z.; Hu, T.T. UAV-based remote sensing using visible and multispectral indices to estimate leaf nitrogen concentration of summer maize. Biosyst. Eng. 2021, 202, 42–54. [CrossRef]

20. Bargiela, G.; Zhou, X.J.; Deo, R.C.; Soar, J.; Whittaker, F.; Wang, H. The modeling of human facial pain intensity based on Temporal Convolutional Networks trained with video frames in HSV color space. Appl. Soft Comput. 2020, 97, 106805. [CrossRef]

21. Hernández-Hernández, J.L.; García-Mateos, G.; González-Esquiva, J.M.; Escarabájal-Henarejos, D.; Ruiz-Canales, A.; Molina-Martínez, J.M. Optimal color space selection method for plant/soil segmentation in agriculture. Comput. Electron. Agric. 2016, 122, 124–132. [CrossRef]
22. Qu, P.X.; Li, T.F.; Li, G.H.; Tian, Z.; Xie, X.W.; Zhao, W.Y.; Pan, X.P.; Zhang, W.D. MCCA-Net: Multi-color convolution and attention stacked network for Underwater image classification. *Cogn. Robot.* **2022**, *2*, 211–221. [CrossRef]

23. Yue, J.B.; Yang, G.J.; Tian, Q.J.; Feng, H.K.; Xu, K.J.; Zhou, C.Q. Estimate of winter-wheat above-ground biomass based on UAV ultra-high- ground-resolution image textures and vegetation indices. *ISPRS J. Photogramm. Remote Sens.* **2019**, *150*, 226–244. [CrossRef]

24. Shanmugasundaram, J.; Raichal, G.; Dency, G.F.; Rajasekaran, P.; Jeevanantham, V. Classification of epileptic seizure using rotation forest ensemble method with 1D-LBP feature extraction. *Mater. Today* **2021**, *57*, 2190–2194. [CrossRef]

25. Oztürk, S.; Akdemir, B. Application of Feature Extraction and Classification Methods for Histopathological Image using GLCM, LBP, LBGLCM, GLRLM and SFTA. *Procedia Comput. Sci.* **2018**, *132*, 40–46. [CrossRef]

26. Ding, X.J.; Yang, F.; Ma, F.M. An efficient model selection for linear discriminant function-based recursive feature elimination. *J. Biomed. Inform.* **2022**, *129*, 104070. [CrossRef] [PubMed]

27. Wang, C.H.; Xiao, Z.Y.; Wu, J.H. Functional connectivity-based classification of autism and control using SVM-RFECV on rs-fMRI data. *Phys. Med.* **2019**, *65*, 99–105. [CrossRef]

28. Djarum, D.H.; Ahmad, Z.; Zhang, J. River Water Quality Prediction in Malaysia Based on Extra Tree Regression Model Coupled with Linear Discriminant Analysis (LDA). *Eur. Symp. Comput. Aided Process Eng.* **2021**, *50*, 1491–1496.

29. Yang, X.; Yang, R.; Ye, Y.; Yuan, Z.R.; Wang, D.Z.; Hua, K.K. Winter wheat SPAD estimation from UAV hyperspectral data using cluster-regression methods. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *105*, 102618. [CrossRef]

30. Xu, J.; Meng, J.H.; Quackenbush, L.J. Use of remote sensing to predict the optimal harvest date of corn. *Field Crops Res.* **2019**, *236*, 1–13. [CrossRef]

31. Islam, M.M.; Matsushita, S.; Noguchi, R.; Ahamed, T. Development of remote sensing-based yield prediction models at the maturity stage of boro rice using parametric and nonparametric approaches. *Remote Sens. Appl. Soc. Environ.* **2021**, *22*, 100494. [CrossRef]

32. Xuan, G.T.; Gao, C.; Shao, Y.Y.; Wang, X.Y.; Wang, Y.X.; Wang, K.L. Maturity determination at harvest and spatial assessment of moisture content in okra using Vis-NIR hyperspectral imaging. *Postharvest Biol. Technol.* **2021**, *180*, 111597. [CrossRef]

33. Pullens, J.W.M.; Srensen, C.A.G.; Olesen, J.E. Temperature-based prediction of harvest date in winter and spring cereals as a basis for assessing viability for growing cover crops. *Field Crops Res.* **2021**, *264*, 108085. [CrossRef]