Estimation of paddy rice maturity using digital imaging

Rui Wang*, Feng Han*, and Wenfu Wu*a,b

*College of Biological & Agricultural Engineering, Jilin University, Changchun, China; bGrain College, Jilin Business and Technology College, Changchun, China

ABSTRACT
Harvest time is an important factor affecting grain yield and postharvest quality. The estimation of crop maturity greatly supports farmers in the optimization of the harvest time. This study proposed a method for predicting paddy rice maturity (J816 and 5Y4 varieties) using color features and the random forest (RF) regression algorithm. Paddy panicle images were obtained using a flatbed scanner during the maturation period. To estimate paddy rice maturity, 22 color features representing the greenness of crop leaves were extracted from the paddy panicle images. Stepwise regression was used to select superior features as inputs to the RF regression model. The coefficient of determination (R²) and root mean square error (RMSE) values of the model were 0.93 and 1.18% for J816 and 0.94 and 1.60% for 5Y4, respectively. The results indicate that the proposed method in this study is a promising technique for the estimation of paddy rice maturity.

Introduction
Crop maturity is an important factor determining the optimum harvest time and affecting grain yield and postharvest quality. A crop produces the maximum dry matter yield at maturity, and then the grains do not continue to grow and only lose moisture. After maturity, the risk of lodging, shattering, and weather changes for crops increases, resulting in reductions in grain quality and yield at harvest. In California, the head rice yield was shown to decrease from 63.8 to 45.8% when the harvest was delayed by 10 days. In Canada, the yield loss of wheat was shown to be as high as 17% due to shattering after delayed harvest. In addition, a delayed harvest leads to an increasing crack rate in rice due to meteorological changes such as rain and dew, affecting the subsequent mechanical drying quality and the taste of milled rice.

Rice is one of the most important cereal crops and supplies nearly 50% of the world’s population as a staple food. The fast increase in the world’s population and the gradual reduction in cropland areas have been raising the demand for foods, especially staple foods such as rice. In general, farmers determine the optimum harvest time in paddy rice mainly through visual observation in the field. Although widely used, this visual method based on personal experience is subjective and highly inconsistent and shows poor reproducibility, resulting in improper harvest times. Kernel dry matter accumulation after anthesis could provide a precise estimation of the optimum harvest time in crops. Crop maturity can be estimated in the maturation period by weighing oven-dried samples and establishing regression models. In addition, grain moisture content as a useful indicator for determining the optimum harvest time has been successfully applied in many grain crops, such as wheat, barley, maize, and rapeseed. However, these methods, which are time-consuming and destructive, in addition to their high costs and labor consumption, cannot be used for real-time applications. Thus, a simple, efficient, and objective method is necessary to determine rice maturity.

CONTACT Wenfu Wu  wwfzlb@126.com  College of Biological & Agricultural Engineering, Jilin University, Changchun 130022, China

Published with license by Taylor & Francis Group, LLC. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
In recent years, nondestructive techniques based on spectral and hyperspectral reflectance have been investigated for estimating the chlorophyll content of plants. These methods were developed to achieve special purposes such as real-time and accurate nutrient status reporting. Because chlorophyll content affects the visual features of leaves, using digital cameras or red, green, and blue (RGB) imaging as low-cost instruments in the visible range has also been used in nitrogen status estimation. On the other hand, mature grains or panicles of paddy rice change color from green to golden brown. Thus, the estimation of paddy rice maturity by its color using digital imaging seems feasible.

Computer vision, as a rapid, low-cost, and high-accuracy technique, has been widely applied in agricultural production. As chlorophyll is the main pigment in leaves and is responsible for leaf greenness, various color indices have been designed to assess plant chlorophyll status at the leaf or canopy scale of the crop. Wang et al. used the green channel minus red channel (GMR or G-R), green channel divided by the red channel (G/R), normalized green index (NGI), normalized red index (NRI), and hue features to estimate biomass, N content, and leaf area index (LAI). They mentioned that the GMR and G/R features had a better correlation to the biomass, N content, and LAI than the other features. Farshad Vesali et al. used 19 color features extracted from corn leaf images to estimate the chlorophyll content of corn leaves. They found that of the other features, the hue feature had the strongest linear relationship with chlorophyll content. In addition to leaf greenness, grain color is an important indicator of crop maturity. Therefore, the color features used to assess plant chlorophyll status would also be useful for estimating paddy maturity.

However, predicting paddy rice maturity by its color using digital imaging is still challenging. There have been few applications of computer vision to determining maturity in paddy rice. The main applications of computer vision in rice have been carried out for classification, the detection of external damage, or the determination of quality parameters. Cho et al. proposed a method to determine paddy maturity by color vision techniques. They used RGB cameras to collect images of panicles at ambient illumination in a laboratory, and then the pictures of the panicles were equally divided into 3 parts: the terminal part, the middle part, and the basal part. Three points were randomly selected in each portion to obtain the mean hue value to determine the paddy maturity. As ripeness gradually progresses from terminal to basal, the distribution of colors at each part of the panicle is uneven. This method of color extraction, in addition to the effects of ambient conditions, results in random and inconsistent maturity prediction. This study aimed to evaluate the effect of various color features from digital images on the estimation of paddy rice maturity in the maturation period.

### Materials and methods

#### Paddy rice samples

Two high-yield rice varieties were used in this study: Jijing816 (J816) and 5You4 (5Y4). J816 is short-grain rice with a 112.8 cm plant height, and 5Y4 is long-grain rice with a 122.8 cm plant height. These two varieties were chosen because they have a similar maturation period of approximately 48 to 55 days after heading (DAH). The experiment was carried out in the experimental fields of the Rice Research Institute of Jilin Province, China during 2019 and 2020. The field size for each variety was 1000 × 300 cm, with a plant spacing of 30 cm. The seed development stages are shown in Table 1.

| Year | Varieties | Sowing date | Heading date | Sampling date | Harvest date |
|------|-----------|-------------|--------------|---------------|--------------|
| 2019 | J816      | 04–25       | 08–05        | 09–20 (46 DAH)| 09–29 (55 DAH) |
|      | 5Y4       |             | 08–03        | 09–20 (48 DAH)| 09–29 (57 DAH) |
| 2020 | J816      | 04–22       | 08–02        | 09–23 (52 DAH)| 09–26 (55 DAH) |
|      | 5Y4       |             | 08–01        | 09–23 (53 DAH)| 09–26 (56 DAH) |
heading date was determined at the time when approximately 50% of the spike had emerged. Paddy panicles were manually harvested every day from the sampling date to the harvest date. At each harvest, enough paddy panicles were randomly selected from healthy plants without any pathogenic symptoms. Then, the paddy panicles were transferred to the laboratory to determine paddy rice maturity and to conduct image acquisition. Image acquisition and the determination of the maturity of the paddy rice were conducted simultaneously to prevent the influence of moisture loss on the paddy color.

**Determination of paddy rice maturity**

To determine paddy rice maturity, paddy samples were threshed by hand from panicles and then dehulled using a laboratory huller (JLGJ-45, China). After this, 20 g sound whole rice samples were collected to determine paddy rice maturity. These rice samples were visually classified into two stages: mature (completely translucent kernel with light brown color) and immature (completely translucent kernel with green color). Finally, rice maturity was determined using Eq. (1).

\[
M = \frac{W_m}{W_m + W_{im}} \times 100\%
\]

where \(M\) is the paddy rice maturity (%), \(W_m\) is the weight of mature kernels (g), and \(W_{im}\) is the weight of immature kernels (g).

**Image acquisition**

To reduce the effects of ambient conditions on imaging, a flatbed scanner (Epson V39, Japan) was used to acquire paddy panicle images. A single panicle was placed on the glass bed of the scanner and then imaged after closing the scanner cover. The scanner interfaced with the computer through a universal serial bus (USB) port. All the images were acquired with a size of \(2550 \times 3509\) pixels at a resolution of 300 dpi and stored in an uncompressed format to avoid introducing any additional artifacts. Color calibration was conducted using a gray card (Kodak R-27, Eastman Kodak Company) before image acquisition on each day.

For each rice variety, 300 images (30 images \(\times\) 10 days) obtained in 2019 were used for model development. These images were randomly split into two groups: 80% (240 images) were selected as the training set to train the model, and the remaining 20% (60 images) were used as the test set to evaluate the performance of the developed model. To validate the robustness of the developed model applied to paddy rice harvested in different seasons, 36 images (9 images \(\times\) 4 days) of each variety acquired in the same fields in 2020 were used in this study. Thus, a total of 600 images were used for model development, and 72 images were used for model validation.

**Image processing**

Image processing aims to isolate the paddy panicle from the original color image. The image processing algorithm was developed using LabView 2016 (National Instruments, USA). The image processing operations of the paddy panicle are shown in Figure 1. After the original panicle image was acquired (Figure 1a), the first step of the image processing operations was thresholding. As the color of the paddy panicle was more saturated than the background, the hue, saturation, and intensity (HSI) color space was suitable for thresholding. From the saturation (S) color space, the threshold of \(S = 25\) was defined in the S-histogram (Figure 2) to extract paddy panicles from the background.

After thresholding, there were still some small particles representing the background in the binary image (Figure 1b). Therefore, a sequence of morphological operations (including opening, particle removal, and border rejection) was applied to filter the noise in the binary image. The opening
operation was a combination of erosion and dilation. Erosion was used to discard the isolated pixels with a higher brightness than the background, followed by dilation to fill the holes or connect cracks inside the objects in the binary image. The particle removal operation was used to filter out particles larger or smaller than a certain size in binary images. Then, the border rejection operation was applied to remove the particles or objects truncated by the boundary in the eroded binary image to enhance the accuracy of object recognition. Finally, the filtered image (Figure 1c) was used as the mask to obtain the isolated color paddy panicle (Figure 1d).

**Feature extraction**

To obtain the color features of the paddy panicles, the RGB and HSI color spaces were used in this study. For each paddy panicle image, the mean and standard deviation of the red (R), green (G), blue (B), hue (H), S, and I) were calculated in the histogram of each color space using Eqs. (2) – (3).

\[
\text{Color}_{\text{mean}} = \frac{\sum_{x=1}^{x} \sum_{y=1}^{y} P_{xy}}{n}
\]

\[
\text{Color}_{\text{std}} = \sqrt{\frac{\sum_{x=1}^{x} \sum_{y=1}^{y} (P_{xy} - \text{Color}_{\text{mean}})^2}{n - 1}}
\]
Table 2. Extracted color features from paddy panicle images.

| No. | Feature Description | Code | Equation |
|-----|---------------------|------|----------|
| 1   | Mean of R           | $R_{mean}$ | Eq. (2) |
| 2   | Mean of G           | $G_{mean}$ |          |
| 3   | Mean of B           | $B_{mean}$ |          |
| 4   | Standard Deviation of R | $R_{std}$ | Eq. (3) |
| 5   | Standard Deviation of G | $G_{std}$ |          |
| 6   | Standard Deviation of B | $B_{std}$ |          |
| 7   | Mean of H           | $H_{mean}$ | Eq. (2) |
| 8   | Mean of S           | $S_{mean}$ |          |
| 9   | Mean of I           | $I_{mean}$ |          |
| 10  | Standard Deviation of H | $H_{std}$ | Eq. (3) |
| 11  | Standard Deviation of S | $S_{std}$ |          |
| 12  | Standard Deviation of I | $I_{std}$ |          |

Combination color features

| No. | Feature Description | Code | Equation |
|-----|---------------------|------|----------|
| 1   | Difference between G and R | GMR | $G_{mean} - R_{mean}$ |
| 2   | G divided by R       | GDR | $G_{mean}/R_{mean}$ |
| 3   | Vegetation index     | VI   | $VI = (G_{mean} - R_{mean})/(G_{mean} + R_{mean})$ |
| 4   | Normalized R         | NR   | $NR = R_{mean}/(R_{mean} + G_{mean} + B_{mean})$ |
| 5   | Normalized G         | NG   | $NG = G_{mean}/(R_{mean} + G_{mean} + B_{mean})$ |
| 6   | Normalized B         | NB   | $NB = B_{mean}/(R_{mean} + G_{mean} + R_{mean})$ |
| 7   | Coefficient of variation of R | CVR | $CVR = R_{mean}/R_{std}$ |
| 8   | Coefficient of variation of G | CVG | $CVG = G_{mean}/G_{std}$ |
| 9   | Coefficient of variation of B | CVB | $CVB = B_{mean}/B_{std}$ |
| 10  | Extreme G            | EXG  | $EXG = 2G_{mean} - R_{mean} - B_{mean}$ |

where $Color_{mean}$ and $Color_{std}$ are the mean and standard deviation of the color values, $n$ is the total number of pixels in the panicle image (Figure 1d), and $p_{xy}$ is the pixel value in the $(x, y)$ position. In addition to these main color features, other combination color features that reflect the greenness of plants proposed in previous studies were calculated. \cite{17,18,22,36,37} Therefore, a total of 22 color features were extracted from each paddy panicle image (Table 2).

**Feature selection**

To improve the performance of the prediction model and to reduce the processing requirements of the computer, it is necessary to select the optimal features that carry the most useful information. In addition, the risk of overfitting problems can be reduced by using limited input variables. Many feature selection methods have been proposed to select optimal features, such as the least absolute shrinkage and selection operator (lasso), decision trees, stepwise regression, and correlation-based feature sections. \cite{18,38,39}

In this study, stepwise regression was used to select the most effective features to develop models for predicting paddy rice maturity. Because stepwise regression can effectively solve collinearity in multiple variables, compared with other methods, it has higher efficiency in selecting features. The basic principle of stepwise regression is to add or remove variables one by one in a regression model based on their statistical significance level until all the independent variables in the model are significant and the other independent variables are insignificant. Stepwise regression was performed using SPSS 19.0 (IBM SPSS Statistics, NY, US).

**Model development**

Random forest (RF) is an ensemble algorithm based on decision trees proposed by Breiman. \cite{40} A decision tree is a nonparametric supervised learning algorithm used for classification and regression. The RF model creates multiple bootstrapped samples from the training set and then builds a number of nonpruned decision trees from each bootstrapped sample set. \cite{41} For regression, the RF model
The Model features model splits used to learn networks of prediction. An RF is a powerful machine learning method that is relatively robust to outliers and can overcome the “black-box” limitations of artificial neural networks. Additionally, the parameterization of RF is very simple and computationally lighter than other machine learning methods (neural networks or support vector machines).

In this study, an RF model was trained with 10-fold cross-validation in Python using the Scikit-Learn package. To obtain the most accurate and robust model, the RandomizedSearchCV method was used for parameter optimization (20 iterations). The optimized parameters and the corresponding searched values were the n_estimators (10 to 500), the max_depth (1 to 30), and the min_samples_split (2 to 10).

The flowchart of model development is shown in Figure 3. Two RF models, the full-features model and the optimal-features model, were developed for each paddy rice variety. The full-features model used all color features extracted from paddy images as the inputs. The optimal-features model used optimal features selected by using stepwise regression as the inputs. The performance of these two models for each paddy rice variety was compared to select a suitable model for predicting maturity.

**Model evaluation**

The evaluation of the RF model for maturity was performed by using the coefficient of determination (R²) and root mean square error (RMSE). Cross-validation was applied to train the developed models. Generally, the larger the R² (close to 1) and the smaller the RMSE (close to 0) are, the better the performance of the developed models. However, what we truly expect is a model that performs well on new samples. For this purpose, the model should learn a ‘universal law’ that is as suitable as possible for all potential samples from the training samples to make correct predictions when performing on new samples.

In this study, a total of 300 images for each paddy rice variety were separated into a training set (240 images) and a test set (60 images). In general, the model gave nearly the same performance on the training set and test set, which means that the model can perform well on new samples. The R² and RMSE can be calculated as...
\[ R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \]  

(4)

\[ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \]  

(5)

where \( y_i \) is the i-th actual maturity, \( \hat{y}_i \) is the i-th predicted maturity, \( \bar{y} \) is the average of actual maturity, and \( n \) is the number of samples.

**Results and discussion**

**Correlation of the color features with maturity**

Figure 4 shows the maturity evolution of paddy rice during the maturation period. In this study, the maturity of the two rice varieties presented a gradually increasing trend during the maturation period. The maturity of J816 was approximately 77% at the harvest date, whereas that of 5Y4 was approximately 57%. Note that a large area of lodging occurred in the field due to the influence of typhoons, resulting in slow growth in the rice, especially for 5Y4, which had a higher plant height than J816.

In the two rice varieties, nine color features, CVR, CVG, CVB, \( R_{\text{std}} \), \( G_{\text{std}} \), \( B_{\text{std}} \), \( I_{\text{std}} \), NG, and EXG, had rather high \( R^2 \) values (> 0.7) (Table 3). Of these features, CVB and \( B_{\text{std}} \) had the best correlation with maturity for J816 (\( R^2 = 0.89 \)) and 5Y4 (\( R^2 = 0.86 \)) rice, respectively. As previously reported, \[22\] most of these greenness features derived from digital camera images were shown to be highly positively correlated with the chlorophyll content of corn leaves. However, in this study, these features were highly negatively correlated with rice maturity. This was expected because when a rice plant grows toward maturity, the grains gradually change from green to yellow due to the loss of nitrogen absorbed by the mature panicle. \[43\] Therefore, our results indicated that these nine color features extracted from each panicle image could be used as predictors for estimating paddy maturity.

![Figure 4. The maturity evolution of paddy rice during the maturation period.](image)
Tewari et al.\textsuperscript{[37]} reported that the mean value of the various image features was found to be best correlated with the greenness of the crop canopy. However, in this study, the mean value of the different color features was poorly fitted to rice maturity because these features were extracted from each segmented image, which contained only the panicle. As ripeness gradually progresses from terminal to basal, the color distribution of the panicle in each part is uneven. In addition, the computation of features in this study used the mean value of all the pixels in the panicle images. Therefore, the correlation between the mean value of colors and maturity was poor, while the standard deviation and coefficient of variation of colors had the strongest relationship with maturity. The GMR (G-R), which others have reported to be valid in the representation of N, was not superior to the other color features in correlating with paddy rice maturity.\textsuperscript{[17]} This was likely due to its limited generality in rice crops, which have unique coexisting impacts from water and soil background and canopy structure.

### Estimation of maturity using all features

Based on all 22 color features extracted from each paddy panicle image, the prediction model was established using the RF algorithm. The performance of the RF model based on all the features for each rice variety is shown in Table 4. As shown in Table 4 and Figure 5, the $R^2$ and RMSE of the full-features model were 0.94 and 1.08% for J816 and 0.95 and 1.43% for 5Y4, respectively. The results indicated that the RF model had a superior performance for the two varieties in the prediction of rice maturity. However, overfitting was present in these models due to the differences in the performance metrics ($R^2$ and RMSE) between the training set and test set, which indicated that the model with all the features may be less stable on new samples.

### Table 3. Relationships of color feature to maturity in paddy rice.

| Feature | Slope | Intercept | $R^2$ | Feature | Slope | Intercept | $R^2$ |
|---------|-------|-----------|------|---------|-------|-----------|------|
| CVB     | −2.33 | 1.55      | 0.89 | $B_{\text{std}}$ | −0.03 | 1.39      | 0.86 |
| CVG     | −4.61 | 1.81      | 0.88 | $I_{\text{std}}$ | −0.04 | 1.47      | 0.82 |
| CVR     | −8.09 | 2.20      | 0.84 | CVB     | −3.59 | 1.89      | 0.80 |
| $B_{\text{std}}$ | −0.02 | 1.24      | 0.80 | CVG     | −5.71 | 2.04      | 0.79 |
| $G_{\text{std}}$ | −0.03 | 1.34      | 0.78 | $G_{\text{std}}$ | −0.04 | 1.48      | 0.79 |
| $I_{\text{std}}$ | −0.03 | 1.29      | 0.78 | CVR     | −10.83 | 2.59      | 0.77 |
| NG      | −23.62 | 8.37     | 0.74 | NG      | −44.39 | 14.84     | 0.72 |
| $R_{\text{std}}$ | −0.04 | 1.64      | 0.74 | EXG     | 0.06  | −0.01     | 0.71 |
| EXG     | 0.03  | 0.48      | 0.72 | $R_{\text{std}}$ | −0.06 | 1.89      | 0.68 |
| $H_{\text{mean}}$ | −0.09 | 2.43      | 0.64 | VI      | 8.04  | −0.73     | 0.64 |
| VI      | 5.67  | −0.23     | 0.58 | GDR     | −5.29 | 4.39      | 0.64 |
| $S_{\text{mean}}$ | −3.81 | 3.46      | 0.58 | NR      | 11.72 | −4.65     | 0.57 |
| $G_{\text{mean}}$ | −0.03 | 1.56      | 0.54 | GMR     | 0.05  | −1.06     | 0.55 |
| NR      | 7.52  | −2.72     | 0.38 | $H_{\text{mean}}$ | −0.16 | 3.48      | 0.55 |
| $I_{\text{mean}}$ | −0.01 | 1.92      | 0.53 | $G_{\text{mean}}$ | −0.02 | 2.42      | 0.53 |
| GMR     | 0.03  | −0.47     | 0.38 | $R_{\text{mean}}$ | −0.02 | 1.87      | 0.51 |
| $B_{\text{mean}}$ | −0.01 | 1.45      | 0.32 | GMR     | −14.17 | 3.86      | 0.46 |
| NB      | 5.18  | 1.86      | 0.10 | $I_{\text{mean}}$ | −0.02 | 2.64      | 0.43 |
| $H_{\text{std}}$ | −0.09 | 0.97      | 0.02 | $S_{\text{mean}}$ | −0.02 | 2.96      | 0.27 |
| $S_{\text{mean}}$ | 0.00  | 0.80      | 0.001| $S_{\text{mean}}$ | 0.01  | −0.62     | 0.11 |

### Table 4. The performance of the RF model with full features for two varieties.

| Varieties | Parameters | n_estimators | min_samples_split | max_depth | Training $R^2$ | RMSE | Test $R^2$ | RMSE |
|-----------|------------|--------------|-------------------|-----------|---------------|------|------------|------|
| J816      | 78         | 3            | 9                 | 0.99      | 0.40          | 0.94 | 1.08       |
| 5Y4       | 119        | 4            | 7                 | 0.99      | 0.73          | 0.95 | 1.43       |
Estimation of maturity using optimal features

Although the RF model based on all the features achieved good performance, many features made the model inefficient and unstable. Therefore, selecting the optimal features to develop the model is necessary. In this study, stepwise regression with a significance level of 0.05 for entry and 0.1 for exclusion was applied to select optimal features as input variables for the RF model. As shown in Table 5, the selected optimal features were CVB, CVG, and H_{std} for J816 and B_{std}, H_{std}, and S_{mean} for 5Y4. The $R^2$ and RMSE of the optimal-features model were 0.93 and 1.18% for J816 and 0.94 and 1.60% for 5Y4, respectively (Figure 6). The RF model with optimal features gave similar performance on the training set and test set (Table 5) and performance equivalent to that of the full-features model (Figure 5). This implied that using the selected optimal features could reduce the overfitting problem without reducing the performance of the model.

### Table 5. The performance of the RF model with optimal features for two varieties.

| Varieties | Features           | Training $R^2$ | RMSE  | Test  $R^2$ | RMSE  |
|-----------|--------------------|----------------|-------|------------|-------|
| J816      | CVB, CVG, and H_{std} | 0.94           | 0.97  | 0.93       | 1.18  |
| 5Y4       | B_{std}, H_{std}, and S_{mean} | 0.95           | 1.38  | 0.94       | 1.60  |

**Figure 5.** Results of the RF model with full features: (a) J816, (b) 5Y4.

**Figure 6.** Results of the RF model with optimal features: (a) J816, (b) 5Y4.

**Estimation of maturity using optimal features**

Although the RF model based on all the features achieved good performance, many features made the model inefficient and unstable. Therefore, selecting the optimal features to develop the model is necessary. In this study, stepwise regression with a significance level of 0.05 for entry and 0.1 for exclusion was applied to select optimal features as input variables for the RF model. As shown in Table 5, the selected optimal features were CVB, CVG, and H_{std} for J816 and B_{std}, H_{std}, and S_{mean} for 5Y4. The $R^2$ and RMSE of the optimal-features model were 0.93 and 1.18% for J816 and 0.94 and 1.60% for 5Y4, respectively (Figure 6). The RF model with optimal features gave similar performance on the training set and test set (Table 5) and performance equivalent to that of the full-features model (Figure 5). This implied that using the selected optimal features could reduce the overfitting problem without reducing the performance of the model.
Validation of the developed models

The best prediction model should have a high $R^2$ value, low RMSE, and the minimum variations in these performance metrics between training and testing. Therefore, it is possible to replace the full-features RF model with the optimal-features RF model to predict rice maturity in this study. For this, the optimal-features RF models of the two varieties were validated using paddy panicle images acquired in 2020. As shown in Figure 7, the performance of the model (the $R^2$ values were 0.92 and 0.96 and the RMSE values were 1.16% and 1.44% for J816 and 5Y4, respectively) showed that the developed models yielded almost the same performance on the test set (collected in 2019) and on the validation set (collected in 2020). The results imply that the proposed method in this study can better predict the rice maturity of J816 and 5Y4 harvested in different seasons.

Farshad Vesali et al.\textsuperscript{[22]} developed an android app for smartphones that estimated the chlorophyll content of a corn leaf using linear (regression) and neural network models based on various color features extracted from digital image. The app estimation compared well with the corresponding soil plant analysis development (SPAD) meter values ($R^2 = 0.88$ and 0.72 and RMSE = 4.03 and 5.96 for the neural network and linear model, respectively). Haw et al.\textsuperscript{[32]} proposed a method using the hue color of florets extracted from digital images to predict paddy maturity. They used a linear regression model with only the hue feature as input and reported that the $R^2$ values for grain hue at the terminal, middle and basal parts were 0.77, 0.91, and 0.96, respectively. As ripeness gradually progresses from terminal to basal, the distribution of colors at each part of the panicle is uneven. This method of color extraction, in addition to the effects of ambient conditions, results in random and inconsistent maturity prediction. Compared to these, our results indicated that the proposed method in this study can be considered a rapid and low-cost alternative for estimating paddy maturity in the field, especially when there is a demand for high availability.

Nondestructive determination of crop maturity is important for acquiring timely and inexpensive information regarding the optimum harvest time and crop management. In this study, we used image processing techniques to predict maturity in two high-yield rice varieties (J816 and 5Y4) to determine the best harvest time. We collected the panicle images with a scanner and extracted various color features to improve the accuracy and stability of the prediction results. Compared to image acquisition with a camera lens, there is no need for any additional attachment (i.e., illumination system and frames) to acquire high-quality panicle images for further analysis. An additional benefit of scanner imaging is that the panicle image has a relatively consistent background, which is convenient for segmenting the background using image processing techniques.

![Figure 7. Results of the RF model applied in 2020: (a) J816, (b) 5Y4.](image-url)
However, the proposed method for predicting maturity in this study was validated only for J816 and 5Y4 paddy rice, and the model developed in this study was not suitable for other paddy varieties. Further research is needed to collect more samples from different growing regions and crop years to develop a general model that could be broadly suitable for paddy rice varieties.

**Conclusion**

In this study, we proposed a method that uses digital imaging and an RF regression model to predict paddy rice maturity during the maturation period. Color images of paddy panicles of two rice varieties (J816 and 5Y4) were acquired by a flatbed scanner during the maturation period. A total of 22 color features were extracted from the paddy panicle images. Stepwise regression was used to select optimal features as the inputs for the RF regression model. The $R^2$ and RMSE of the RF model with optimal features were 0.93 and 1.18% for J816 and 0.94 and 1.60% for 5Y4, respectively. Compared with the full-features model, the optimal-features RF model yielded equivalent performance without overfitting the data. The validation of the optimal-features model confirmed that it is robust for J816 and 5Y4 rice varieties harvested in different seasons. The digital imaging estimation compared well with the corresponding observation maturity results (the $R^2$ and RMSE were 0.92 and 1.16% for J816 and 0.96 and 1.44% for 5Y4).

The proposed method in this study can be a practical low-cost alternative for measuring paddy maturity. In addition to the lower cost, this method has several other advantages, such as rapid and robust prediction without sample preprocessing. In addition, in comparison with conventional computer vision systems, which usually include an illumination system and image acquisition box, the proposed method in this study requires only a flatbed scanner, which provides technical support for the harvesting-storage management of high-quality rice in field conditions. However, further research on the method is required to validate the performance and robustness of various rice varieties from different growing regions and crop years.

**Conflicts of interest**

The authors declare no conflicts of interest

**Funding**

This work was supported by the Ministry of Science and Technology of the People’s Republic of China (grant number: 2016YFD0401001); the Ministry of Science and Technology of China [2016YFD0401001];

**References**

[1] Calderini, D. F.; Abeledo, L. G.; Slafer, G. A. Physiological Maturity in Wheat Based on Kernel Water and Dry Matter. *Agron. J.* 2000, 92(5), 895–901. DOI: 10.2134/agronj2000.925895x.

[2] Farrer, D.; Weisz, R.; Heiniger, R.; Murphy, J. P.; Pate, M. H. Delayed Harvest Effect on Soft Red Winter Wheat in the Southeastern USA. *Agron. J.* 2006, 98(3), 588–595. DOI: 10.2134/agronj2005.0211.

[3] Thompson, J. F.; Mutters, R. G. Effect of Weather and Rice Moisture at Harvest on Milling Quality of California Medium-grain Rice. *Trans. ASABE.* 2006, 49(2), 435–440. DOI: 10.13031/2013.20392.

[4] Clarke, J. M. Effect of Delayed Harvest on Shattering Losses in Oats, Barley and Wheat. *Can. J. Plant Sci.* 1981, 61 (1), 25–28. DOI: 10.4141/cjps81-004.

[5] Xu, Y.; Ying, Y.; Ouyang, S.; Duan, X.; Sun, H.; Jiang, S.; Sun, S.; Bao, J. Factors Affecting Sensory Quality of Cooked Japonica Rice. *Rice Sci.* 2018, 25(6), 330–339.

[6] Mishra, R.; Rao, G. J. N.; Rao, R. N.; Kaushal, P. Development and Characterization of Elite Doubled Haploid Lines from Two Indica Rice Hybrids. *Rice Sci.* 2015, 22(6), 290–299. DOI: 10.1016/j.rsci.2015.07.002.

[7] Hong, F.; Dong-Dong, C.; Wei-Min, H.; Ya-Jing, G.; Yu-Ying, F. Studies on Optimum Harvest Time for Hybrid Rice Seed. *J. Sci. Food Agric.* 2016, 97, 1124-1133.
[8] Wang, X.; Zheng, H.; Tang, Q. Early Harvesting Improves Seed Vigour of Hybrid Rice Seeds. Sci. Rep. 2018, 8(1), 11092. DOI: 10.1038/s41598-018-29021-5.

[9] Pepler, S.; Gooding, M. J.; Ellis, R. H. Modelling Simultaneously Water Content and Dry Matter Dynamics of Wheat Grains. Field Crops Res. 2006, 95(1), 49–63. DOI: 10.1016/j.fcr.2005.02.001.

[10] Alvarez Prado, S.; Gallardo, J. M.; Serrago, R. A.; Kruk, B. C.; Miralles, D. J. Comparative Behavior of Wheat and barley Associated with Field Release and Grain Weight Determination. Field Crops Res. 2013, 144, 28–33. DOI: 10.1016/j.fcr.2012.12.018.

[11] Parvez, M. R.; Hurburgh, C. R.; Hanna, H. M.; Licht, M. A. Dynamics of Corn Dry Matter Content and Grain Quality after Physiological Maturity. Agron. J. 2020, 112, 98–1011.

[12] Menendez, Y. C.; Botto, J. F.; Gomez, N. V.; Miralles, D. J.; Rondonini, D. P. Physiological Maturity as a Function of Seed and Pod Water Concentration in Spring Rapeseed (Brassica Napus L.). Field Crops Res. 2019, 231, 1–9. DOI: 10.1016/j.fcr.2018.11.002.

[13] Sonobe, R.; Yamashita, H.; Mihara, H.; Morita, A.; Ikka, T. Hyperspectral Reflectance Sensing for Quantifying Leaf Chlorophyll Content in Wasabi Leaves Using Spectral Pre-processing Techniques and Machine Learning Algorithms. Int. J. Remote Sens. 2021, 42(4), 1311–1329. DOI: 10.1080/01431161.2020.1826065.

[14] Amirruddin, A. D.; Muharam, F. M.; Ismail, M. H.; Ismail, M. F.; Tan, N. P.; Karam, D. S. Hyperspectral Remote Sensing for Assessment of Chlorophyll Sufficiency Levels in Mature Oil Palm (Elaeis Guineensis) Based on Frond Numbers: Analysis of Decision Tree and Random Forest. Comput. Electron. Agric. 2020, 169, 11. DOI: 10.1016/j.compag.2020.105221.

[15] Eshkabilov, S.; Lee, A.; Sun, X.; Lee, C. W.; Simsek, H. Hyperspectral Imaging Techniques for Rapid Detection of Nutrient Content of Hydroponically Grown Lettuce Cultivars. Comput. Electron. Agric. 2021, 181, 105968. DOI: 10.1016/j.compag.2020.105968.

[16] Gupta, S. D.; Ibaraki, Y.; Pattanayak, A. K. Development of a Digital Image Analysis Method for Real-time Estimation of Chlorophyll Content in Micropropagated Potato Plants. Plant Biotechnol. Rep. 2013, 7, 97-97.

[17] Wang, Y.; Wang, D.; Zhang, G.; Wang, J. Estimating Nitrogen Status of Rice Using the Image Segmentation of G-R Thresholding Method. Field Crops Res. 2013, 149(149), 33–39. DOI: 10.1016/j.fcr.2013.04.007.

[18] Lee, K. J.; Lee, B. W. Estimation of Rice Growth and Nitrogen Nutrition Status Using Color Digital Camera Image Analysis. Eur. J. Agron. 2013, 48, 57–65. DOI: 10.1016/j.eja.2013.02.011.

[19] Santos Gomes, J. F. S.; Leta, F. R. Applications of Computer Vision Techniques in the Agriculture and Food Industry: A Review. Eur. Food Res. Technol. 2012, 235(6), 989–1000. DOI: 10.1007/s00217-012-1844-2.

[20] Zareiforoush, H.; Minaei, S.; Alizadeh, M. R.; Banakar, A. Potential Applications of Computer Vision in Quality Inspection of Rice: A Review. Food Eng. Rev. 2015, 7(3), 321–345. DOI: 10.12938/12393-014-9101-z.

[21] Patricio, D. I.; Rieder, R. Computer Vision and Artificial Intelligence in Precision Agriculture for Grain Crops: A Systematic Review. Comput. Electron Agric. 2018, 153, 69–81. DOI: 10.1016/j.compag.2018.08.001.

[22] Vesali, F.; Omid, M.; Kaleita, A.; Mobli, H. Development of an Android App to Estimate Chlorophyll Content of Corn Leaves Based on Contact Imaging. Comput. Electron. Agric. 2015, 116, 211–220. DOI: 10.1016/j.compag.2015.06.012.

[23] Lin, P.; Li, X. L.; Chen, Y. M.; He, Y. A Deep Convolutional Neural Network Architecture for Boosting Image Discrimination Accuracy of Rice Species. Food Bioprocess. Technol. 2018, 11(4), 765–773. DOI: 10.1007/s11947-017-2050-9.

[24] Kiratitranapruk, K.; Tenniranrat, P.; Sinhupinyo, W.; Prempree, P.; Chaitavon, K.; Pornthearaphat, S.; Prasertsak, A. Development of Paddy Rice Seed Classification Process Using Machine Learning Techniques for Automatic Grading Machine. J. Sens. 2020. DOI: 10.1155/2020/7041310.

[25] Wang, L.; Liu, M.; Liu, X.; Wu, L.; Wan, P.; Wu, C. Pretrained Convolutional Neural Network for Classifying Rice-cropping Systems Based on Spatial and Spectral Trajectories of Sentinel-2 Time Series. J. Appl. Remote Sens. 2020, 14, 1. DOI: 10.1117/1.jrs.14.014506.

[26] Sun, K.; Wang, Z.; Tu, K.; Wang, S.; Pan, L. Recognition of Mould Colony on Unhulled Paddy Based on Computer Vision Using Conventional Machine-learning and Deep Learning Techniques. Sci. Rep. 2016, 6(1). DOI: 10.1038/srep37994.

[27] Chen, S.; Xiong, J.; Guo, W.; Bu, R.; Zheng, Z.; Chen, Y.; Yang, Z.; Lin, R. Colored Rice Quality Inspection System Using Machine Vision. J. Cereal. Sci. 2019, 88, 87–95. DOI: 10.1016/j.jcres.2019.05.010.

[28] Chen, J.; Lian, Y.; Zou, R.; Zhang, S.; Ning, X.; Han, M. Real-time Grain Breakage Sensing for Rice Combine Harvesters Using Machine Vision Technology. Int. J. Agric. Biol. Eng. 2020, 13(3), 194–199. DOI: 10.25165/j.ijabe.20201303.5478.

[29] Payman, S. H.; Bakhtshipour, A.; Zareiforoush, H. Development of an Expert Vision-based System for Inspecting Rice Quality Indices. Qual. Assur. Saf. Crops Food. 2018, 10(1), 103–114. DOI: 10.3920/qas2017.1109.

[30] Chen, J.; Lian, Y.; Li, Y. Real-time Grain Impurity Sensing for Rice Combine Harvesters Using Image Processing and Decision-tree Algorithm. Comput. Electron. Agric. 2020, 175, 105591. DOI: 10.1016/j.compag.2020.105591.

[31] Singh, S. K.; Vidyarthi, S. K.; Tiwari, R. Machine Learnt Image Processing to Predict Weight and Size of Rice Kernels. J. Food Eng. 2020, 274. DOI: 10.1016/j.jfoodeng.2019.109828.
[32] R S. Colour Vision to Determine Paddy Maturity. Int. J. Agric. Biol. Eng. 2014, 7(5), 55–63. DOI: 10.3965/j.ijabe.20140705.006.

[33] Diaz, E. O.; Kawamura, S.; Koseki, S. Effect of Thickness and Maturity on Protein Content of Japonica Brown Rice Collected during Postharvest Processing. Biosyst. Eng. 2019, 183, 160–169. DOI: 10.1016/j.biosystemseng.2019.04.023.

[34] Ramirez-Paredes, J.-P.; Hernandez-Belmonte, U.-H. Visual Quality Assessment of Malting Barley Using Color, Shape and Texture Descriptors. Comput. Electron. Agric. 2020, 168, 105110. DOI: 10.1016/j.compag.2019.105110.

[35] Garrido-Novell, C.; Pérez-Marín, D.; Amigo, J. M.; Fernández-Novales, J.; Guerrero, J. E.; Garrido-Varo, A. Grading and Color Evolution of Apples Using RGB and Hyperspectral Imaging Vision Camera. J. Food Eng. 2012, 113(2), 281–288. DOI: 10.1016/j.jfoodeng.2012.05.038.

[36] Woebbecke, D. M.; Meyer, G. E.; Von Bargen, K.; Mortensen, D. A. Color Indices for Weed Identification under Various Soil, Residue, and Lighting Conditions. Trans. ASAE. 1995, 38(1), 259–269. DOI: 10.13031/2013.27838.

[37] Tewari, V. K.; Arudra, A. K.; Kumar, S. P.; Pandey, V.; Chandel, N. S. Estimation of Plant Nitrogen Content Using Digital Image Processing. Agric. Eng. Int. CIGR e-j. 2013, 15(2), 78–86.

[38] Mollazade, K.; Omid, M.; Arefi, A. Comparing Data Mining Classifiers for Grading Raisins Based on Visual Features. Comput. Electron. Agric. 2012, 84(0), 124–131. DOI: 10.1016/j.compag.2012.03.004.

[39] Teimouri, N.; Omid, M.; Mollazade, K.; Rajabipour, A. A Novel Artificial Neural Networks Assisted Segmentation Algorithm for Discriminating Almond Nut and Shell from Background and Shadow. Comput. Electron. Agric. 2014, 105, 34–43. DOI: 10.1016/j.compag.2014.04.008.

[40] Breiman, L. Random Forests. Mach. Learn. 2001, 45(1), 5–32. DOI: 10.1023/A:1010933404324.

[41] Guo, L.; Chehata, N.; Mallet, C.; Boukir, S.; Eshkabilov, S.; Lee, A.; Sun, X.; Lee, C. W.; Simsek, H. Relevance of Airborne Lidar and Multispectral Image Data for Urban Scene Classification Using Random Forests. Isprs J. Photogramm. Remote Sens. 2011, 66(1), 56–66. DOI: 10.1016/j.isprsjprs.2010.08.007.

[42] Rahimi-Ajdadi, F.; Abbaspour-Gilandeh, Y.; Mollazade, K.; Hasanzadeh, R. P. R. Development of a Novel Machine Vision Procedure for Rapid and Non-contact Measurement of Soil Moisture Content. Measurement. 2018, 121, 179–189. DOI: 10.1016/j.measurement.2018.02.060.

[43] Laza, R. C.; Parao, F. T.; Akita, S. Growth and Carbon Balance of a Rice Canopy Grown in Temperature-controlled Conditions. Philipp. J. Crop Sci. (Philipp.). 1989, 14(2), 69-75.