Article

Not peer-reviewed version

A non-destructive monitoring method for chlorophyll content of potato plants based on feature reconstruction algorithm

Yaohua Hu *, Huanbo Yang , Xiaoyi Shi , Kaili Zhang , Taifeng Guo

Posted Date: 3 April 2024

doi: 10.20944/preprints202404.0311.v1

Keywords: chlorophyll content estimation; UAV- based multispectral images; NDRE; PCA3 and INDRE.

Article

A Non-Destructive Monitoring Method for Chlorophyll Content of Potato Plants Based on Feature Reconstruction Algorithm

Yaohua Hu [1,*,†], Huanbo Yang [2,†], Xiaoyi Shi [1], Kaili Zhang [2] and Taifeng Guo [2]

[1] Zhejiang A&F University, Hangzhou 311300, China; shixiaoyi@stu.zafu.edu.cn
[2] Northwest A&F University, Yangling 712100, China; yanghuanbo@nwsuaf.edu.cn (H.Y.); kelly999@nwafu.edu.cn (K.Z.); gtf2021@nwafu.edu.cn (T.G.)
* Correspondence: huyaohua@zafu.edu.cn
† These authors contributed equally to this work.

Abstract: Using UAV-based multispectral images for quickly and accurately monitoring chlorophyll content is critical for field management and yield estimation. However, the lower model accuracy and the poor robustness of the estimation models are still preventing the widespread application of UAV-based multispectral images. We carried out two field trials at various experimental sites to further enhance the precision and applicability of the model used to estimate the chlorophyll content of potato plants. Firstly, the texture features and vegetation indices derived from multispectral images were screened using the Pearson correlation coefficient method, and Normalized difference red edge (NDRE) performed the best over the two growth periods. Secondly, principal component analysis (PCA) was applied to recombine five bands of multispectral images, and third PCA results (PCA3) was selected to combined with NDRE according to the construction principle of NDRE, and the newly constructed parameter was named improved NDRE (INDRE). Finally, INDRE was used to establish a chlorophyll content estimation model of potato plants, and compared with some traditional parameters. The results demonstrated that the INDRE had the maximum accuracy (R^2 = 0.7865, RMSE = 2.1378), and corresponding R^2 increased by 0.1481 and RMSE decreased by 1.2994 than NDRE. Additionally, the model was validated using independent data from Experiment 2, and INDRE considerably increased estimation accuracy compared to other factors. In conclusions, the INDRE suggested in this study significantly enhances the accuracy and applicability of the chlorophyll content inversion model and can serve as an additional reference for fertilization management.

Keywords: chlorophyll content estimation; UAV- based multispectral images; NDRE; PCA3 and INDRE

1. Introduction

Chlorophyll is a key pigment that collects solar radiation and converts chemical energy [1], and photosynthesis is a crucial indicator of the growth and development of potato plants [2]. Chlorophyll content has been proven to be closely related to nitrogen (N) concentration [3,4], which is used to quantify crop photosynthetic capacity and monitor the nutritional status of crops or to guide field management such as irrigation [5], fertilization [6], and pesticide application [7]. Therefore, rapid and accuracy monitoring of crop chlorophyll content is of great significance for increasing crop yield.

Currently, the most accurate method for measuring crop chlorophyll content is by stoichiometric [6]. However, destruction sampling is laborious and inefficient for measurement of chlorophyll content in large area crops [1,8]. Crop phenotypic information collection within field scale has become a bottleneck in precision agriculture [9,10]. In recent years, remote sensing technology gradually place in the forefront position of modern crop growth monitoring due to its advantages of quickly,

 doi:10.20944/preprints202404.0311.v1

accuracy, and non-destruction [11–13]. SPAD meter is widely adopted to measure chlorophyll content of crops [14,15], and the measurement results reflect the actual chlorophyll content [16], and can be used for nitrogen (N) concentration diagnosis [17,18]. However, the time-consuming, high labor intensity confine its further application in large areas detection. In practical applications, SPAD instruments are often used instead of stoichiometry to verify the performance of other remote sensing platforms suitable for field scale.

The generation of UAV remote sensing platforms, spanning from multi-spectral to hyperspectral data, to estimate the chlorophyll content of crops has considerably increased. Compared with SPAD meter, UAV remote sensing platform is no longer limited by field area, and can significantly improve data collection efficiency. A variety of vegetation indices (VIs) calculated from UAV-based multispectral images are commonly used to assess chlorophyll content due to the advantages of its straightforward and effective [19]. In the study of Yang et al., six VIs were used to construct chlorophyll retrieval models for potato plants during two growth periods, and the results showed that stacking algorithm yield the maximum prediction accuracy (R^2 = 0.672, RMSE = 0.537) [20]. Moreover, Singhal et al. highlighted the significance correlation between red band and chlorophyll content in maize plants, and when comparing the performance of several machine learning algorithms, Kernel-Ridge obtained the most robust estimation results [21]. Although, several positive progresses have been made relied on multispectral VIs, problems retained. One is, when the crop biomass content reaches a critical value, multispectral VIs prone to lose sensitivity, and this phenomenon is regarded as a saturation problem. In addition, the existing multispectral VIs are mainly constructed on the basis of spectral differences between background and crop canopy, which often lead to lower accuracy in estimating chlorophyll content.

UAV based-hyperspectral images have gained great popularity in chlorophyll content prediction [8], owning to its capacity to alleviate the saturation problem [22]. Hyperspectral images provide more band ranges and can characteristic crop canopy spectrum in great detail than multispectral images [12]. Sun et al. analyzed the corresponding relationship between the hyperspectral information and the chlorophyll content of corn under different lodging levels, and developed a model for estimating chlorophyll content using complex processing techniques [23]. Among these variables, VIs achieved the highest chlorophyll content retrieval accuracy compared with other common mathematical transformation processing methods for maize under lodging stress. However, the autocorrelation of hyperspectral data and the selection of sensitive bands have always been important issues in the process of hyperspectral data [24]. In order to solve these problems, Zhang et al. created a new feature selection method based on the 2DCOS, and introduced a transfer learning algorithm to establish a regression model for reducing complexity and improving robustness. The models of validation results have well accuracy and robustness [22]. However, high cost for hyperspectral sensor [25] and complex processing methods in hyperspectral data confine the further application in chlorophyll content prediction compared to UAV multispectral remote sensing platform.

It is imperative to fully excavate the information from multispectral images, and further improve the accuracy of chlorophyll content inversion. Texture features reflect the canopy structure characteristics of crops, and combining them with multispectral VIs can alleviate the saturation problem with a certain extent, thereby improving the model accuracy [12,20]. But this improvement in accuracy is limited. Moreover, Yang et al. collected images of potato plant canopy taken with RGB cameras and explored relationship between the bands, further constructed the RGFI [20]. RGFI produced the highest accuracy in chlorophyll content estimation compared to other traditional vegetation indices derived from RGB cameras. At the same time, Yang et al. obtained the highest accuracy of fractional vegetation cover (FVC) estimation based on the fitting results in the red and green bands [26]. There are fewer wavebands in RGB images than multispectral images, which lead to lower model accuracy and robustness, thus, UAV RGB platform is rarely used for chlorophyll content estimation of field crops. Principal component analysis (PCA), as a classic feature dimensionality reduction method by finding the direction with the highest variance in the data, have been widely used in hyperspectral image processing [27,28] However, so far, few studies have

 doi:10.20944/preprints202404.0311.v1

applied PCA to quantify crop chlorophyll content based on UAV- multispectral images. Moreover, inversion models based solely on PCA results for crop phenotype information often perform poorly. We hypothesize that the combination of PCA results with VIs can improve the accuracy of chlorophyll content estimation model, while the combination regulation of multispectral VIs and PCA results has not yet been investigated.

The objective in this study was to explore use of combination of PCA results with VIs in UAV-multispectral bands to enhance the multispectral VIs, thereby improved the estimation accuracy of chlorophyll content. The detailed steps were to (1) screen published multispectral VIs and texture features at every stage of potato growth; (2) combine the PCA results with multispectral index that has the highest correlation with chlorophyll content (hereafter referred to INDRE); and (3) compare the predictive performance of traditional VIs and texture features after Pearson correlation coefficient method screening with INDRE, and demonstrate if the combination of PCA results with VI could further improve the estimation accuracy.

2. Materials and Methods

2.1. Study Area

In 2022, potato plants field experiments were conducted, named Experiment 1 (Exp. 1) and Experiment 2 (Exp. 2).

Exp. 1 was located in Ningtiaoliang Town, Yulin City, Shaanxi Province (37°33'55. 57 " N, 108°22'17.46" E) (Figure 1(a)). The climate belongs to emi-arid continental monsoon with annual average precipitation and annual average sunshine close to 395.4 millimeters and 2768.2 hours, respectively. The local type of soil is mainly sandy, with a large temperature difference between day and night, potatoes are widely planted locally. Exp. 2 was conducted in Letangbao Township, Yulin City, Shaanxi Province (37°14'26.09" N, 110°12'38.43" E) (Figure 1(b)). The land is mainly composed of mountains and hill, and annual average precipitation is less approximately 500 mm.

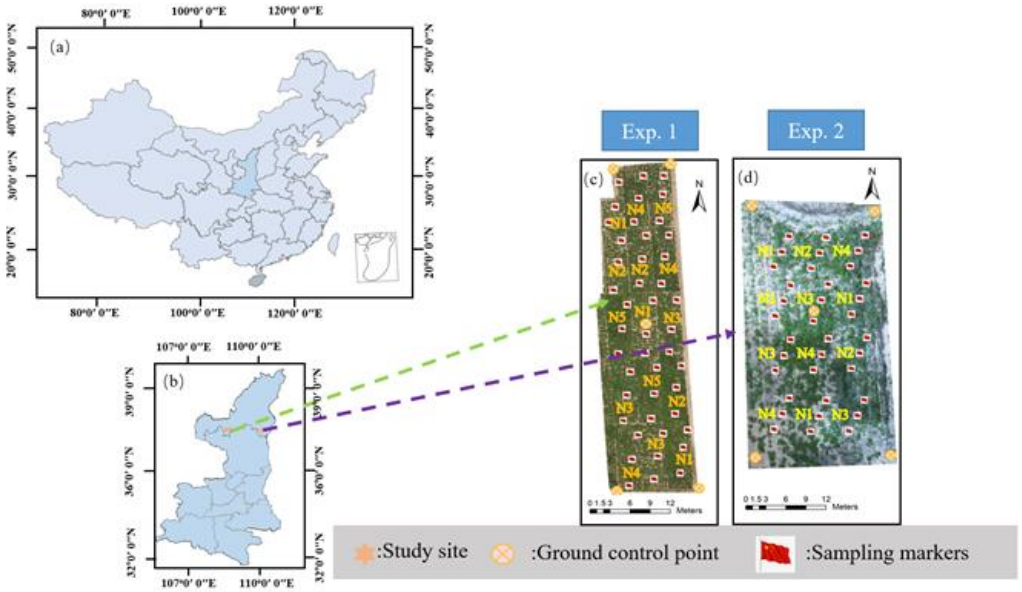


Figure 1. Location and division of the two research fields and UAV-based field observations: (a) location of the Shaanxi province in China, (b) location of the research field in Shaanxi province, (c) and (d) shows the aerial view of Exp. 1 and Exp. 2 respectively. Moreover, the detailed planning of the ground control points and sampling plots are displayed.

2.2. Experiment Description

The potato varieties tested for Exp. 1 were XISEN No. 6 (P 1), No. 226 (P 2) and V 7(P 3) first-class virus-free seed potatoes. On flat ground, potatoes were planted with a 0.6 m plant spacing and

 doi:10.20944/preprints202404.0311.v1

a 0.5 m row spacing. Drip irrigation systems were installed on one side of potatoes after planting to improve the utilization efficiency of water. The Exp. 1 was setup with three replicate plots and each of the five treatments conformed to a complete random block design (Figure 1(c)). The total area of the experimental field was approximately 1700 m2, the length and width of each treated plot were 14 m and 8 m respectively. The entire experimental field was fully irrigated, but differentiated by N fertilizer application rate, and divided into 5 N fertilizer gradients and 3 replicates including: A1: (N fertigation 0 kg ha−1), A 2: (e.g. 20 kg ha−1 for basic N fertigation, 15 kg ha−1 for the 1st N fertigation , 15 kg ha−1 for the 2st N fertigation), A 3: (e.g. 40 kg ha−1 for basic N, 30 kg ha−1 for the 1st N fertigation , 30 kg ha−1 for the 2st N fertigation), A 4: (e.g. 60 kg ha−1 for basic N, 60 kg ha−1 for the 1st N fertigation, 60 kg ha−1 for the 2st N fertigation), A 5: (e.g. 80 kg ha−1 for basic N, 120 kg ha−1 for the 1st N fertigation , 120 kg ha−1 for the 2st N fertigation). Potatoes planting, emergence and harvest dates are 1 May, 29 May and 27 August, 2022, respectively with a 90-day lifespan.

The potato varieties in Exp. 2 were P 1 and P 3 first-class virus-free seed potatoes. Potatoes were planted by artificial ridging with row spacing of 0.6 m and 0.5 m of plant spacing. As similar to Exp. 1, N fertilizer was controlled as variable with three replicate plots of four treatments including: N 1: (N fertigation 0 kg ha−1), N 2: (N fertigation 80 kg ha−1), N 3: (N fertigation 160 kg ha−1), N 4: (N fertigation 240 kg ha−1), and each treatment laid out a complete random block design. The area of Exp. 2 was approximately 1200 m2, and the dimensions of each plot was 12 m × 8 m. Since Exp. 2 was located between two peaks with an altitude of 150 meters, applying N fertilizer once was conducted during the entire growth period of potato. As the potato varieties in Exp. 1 included that of in Exp. 2, the growth characteristics of potato in Exp. 2 were basically the same as those in Exp. 1. Affected by the rainstorm, we only collected data for one period of potato in Exp. 2.

2.3. Chlorophyll Content Measurement

SPAD reading has been proved to have a significant linear relationship with the chlorophyll content measured by the stoichiometric method, and which can be used instead of stoichiometry. Drawing on previous methods for measuring chlorophyll content [20], we randomly selected three potato plants with similar growth in each plot for measurement. In Exp. 1, measurements were separately taken 60 and 75 days after potato planting, which named separately as V 1 and V 2 stage, and the measuring area should avoid the leaf veins as much as possible. 9 leaves from each plant were select, each leaf was measured 4 times and the corresponding average value served as a measurement of chlorophyll content. The chlorophyll content measurement method in Exp. 2 was consistent with Exp. 1, and the sampling time was 60 days after potato planting.

2.4. UAV -Based Multispectral Images Collection and Preprocessing

On the same day as chlorophyll content measurement, UAV-based multispectral images Acquisition of. The UAV used was a DJI Phantom 4 Multispectral (P4M) UAV (SZ DJI Technology Co., Ltd., Shenzhen, China) with maximum horizontal flight speed in positioning state of 50 km h-1, and maximum flying altitude of 6000 m. The P4M is a multispectral sensor that can take pictures in 5 bands, including the visible and infrared spectrum. All channels are arranged in an array with effective pixels of 2.00 million. Moreover, Real-time kinematic (RTK) system is integrated P4M UAV, which can achieve centimeter level positioning, greatly facilitating accurate management.

The collection of UAV based-multispectral images was carried out at approximately 11:30 am under sunny and windless weather. Before performing a flight mission, the multispectral imagery of two radiometric scales with spectral reflectance of 25% and 50% (GZ Changhui Electronic Technology Co., Ltd, China) were firstly obtained at altitude of 1 m for UAV flight. DJI GS Pro (SZ DJI Technology Co., Ltd., Shenzhen, China) was applied to plan route missions, an application that run on iPad that guided the serpentine image acquisition manner. Detailed UAV platform and multispectral camera information were given in Table 1. According to the suggestion by González-Jaramillo et al. [29], five GCPs were regularly placed throughout the experimental plot in a visible area for georeferencing position of research objects. The P4M UAV flied at an altitude of 30 m and a speed of 3 m/s, with the camera vertically downwards to obtain multispectral images of the potato plants canopy, and

 doi:10.20944/preprints202404.0311.v1

forward and side overlap both are 85%. The time interval for taking photos was 3s, and the route area was designed to be larger than the experimental plot area to ensure sufficient number of point clouds in the boundary area. In addition, Pix4Dmapper software (Pix4D Inc., Lausanne, Switzerland) was used to performed image mosaic processing, included: (1) Inputting all images and performing initialization processing; (2) Importing 5 GCPs into the software, and then conducting one click automation processing; (3) exporting Digital Orthophoto Map (DOM) and Digital Surface Model (DSM).

Table 1. Specific parameter list of P4M UAV.

UAV		**Camera**	
Parameters	Values	Parameters	Values
Product type	Quadcopter	Color output	Global shutter, and all spectral bands aligned
Longest flight time /min	27	Focal length/mm	5.74
Maximum takeoff weight /kg	1.487	Field of view/ (◦)	62.7
Operating temperature/℃	0-40	Pixels	1600×1300
Digital communication distance/km	7	Wave length/mm	200-800
Maximum withstand wind speed(m/s)	8	Capture rate (time/s)	1

2.5. Methodology

The main process consists of four parts (Figure 2). First, the DOM of experimental area was cropped, and calculated the common 8 multispectral VIs and 8 texture features. Second, VIs and texture features were screened, and pick out the VI that has the highest correlation by using Pearson correlation coefficient method. Third, PCA was conducted on the five bands of multispectral images for dimensionality reduction, and PCA result was selected to combine with VI that have well correlation during the two growth periods for further construct a new feature. Fourth, chlorophyll content estimation models were established separately at each growth and the whole growth period, and then evaluated the model accuracy through the measurements. Finally, the chlorophyll content obtained by Exp. 2 was used to additionally evaluate model accuracy and universality.

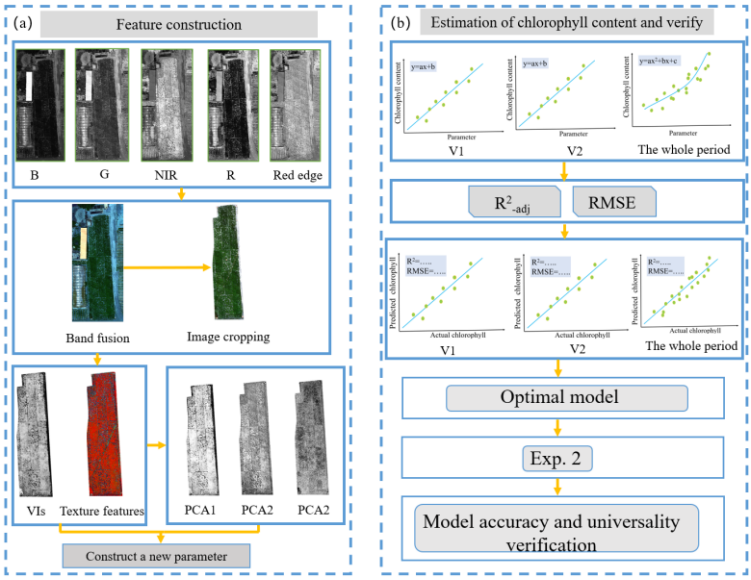


Figure 2. Process for (a): UAV-based multispectral images preprocessing and feature construction, (b): The establishment of chlorophyll content inversion model and verify.

 doi:10.20944/preprints202404.0311.v1

2.5.1. The Calculation of VIs and Texture Features

The physiological characteristics of crops result in relatively low reflectivity of green leaves at 400-490 nm (blue band) and 650-700 nm (red band), while slightly higher reflectivity occurs at 490-600 nm (green band) [1]. Moreover, crop reflectance has increased rapidly from 700 nm. A series of mathematical transformation method is additionally introduced to expand this spectral difference, thereby constructing various multispectral VIs, the widely used VIs mainly include Normalized Difference Vegetation Index (NDVI) [30], soil-adjusted vegetation index (SAVI) [31], optimized soil-adjusted vegetation index (OSAVI) [32], Modified simple ratio (MSR) [33], Ratio Vegetation Index (RVI) [34], Land Cover Index (LCI) [35], Green NDVI (GNDVI) [36], Normalized Difference Red Edge Index (NDRE) [37], Green Chlorophyll Index (GCI) [6] and Atmospherically Resistant Vegetation Index (ARVI) [1]. Gray-level cooccurrence matrix was applied to extract the classic texture features such as mean, correlation(cor), dissimilarity(dis), contrast(con), homogeneity (hom), entropy(ent), angular second moment(sm) and variance(var). A total of 40 texture features were extracted in five bands.

2.5.2. Screen of Characteristic Parameters

Pearson correlation coefficient method, as a typical evaluation variable correlation degree algorithm can screen variables based on standardizing covariance [38]. The strength of a correlation between two variables is measured using Pearson's correlation coefficient (r), which ranges in value from - 1 to 1. Positive and negative values of the r value represent positive and negative correlations, respectively, and the higher the absolute value of the r value, the stronger the correlation. The two growth periods, parameters with better correlation to chlorophyll content were used for subsequent analysis, such as feature construction and establishment of chlorophyll content inversion model.

2.5.3. Characteristic Parameter Construction

PCA usually serves as an important tool for feature dimensionality reduction, which can preserve as much effective parts of the data as possible. At the same time, VIs have great potential to expand the difference between observation and background, but cannot retain sufficient information from the original data [39]. If VIs with well correlation to chlorophyll content and PCA results are combined, canopy phenotype information of crops can be characterized in great detail, and sufficient information from the original data can be retain as much as possible. In this study, the relationship between PCA results and chlorophyll content was described, and the linear relationship between VI and PCA results was statistically analyzed. In addition, the construction principle of the VI with the best correlation to chlorophyll content was applied for the combination of VI and PCA results.

2.5.4. Accuracy Evaluation

The datasets obtained in Exp.1 were used for modeling and evaluation, and training sets and validation sets was randomly divided at a ratio of 7:3 Coefficient of determination (R^2), root mean square error (RMSE), as a typical model evaluation indicator, were applied to verify the accuracy of the regression model. Moreover, applying the optimal model of chlorophyll content to the data obtained in Exp.2 to test the universality and accuracy.

3. Results

3.1. Feature Filtering Results

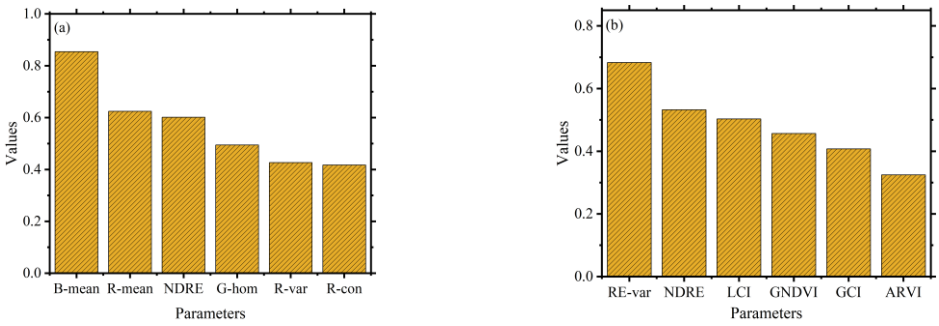
For VIs, after PCA screening, it was found NDRE had the best performed at V1 stage with correlation coefficient value of 0.6011, followed by GOSAVI (r=0.3447) and LCI (r=0.3277). Moreover, the correlation coefficient of the remaining VIs was less than 0.30. At V2 stage, NDRE also performed the best with r value of 0.5323, followed by LCI (r=0.5029), GNDVI (r= 0.4564) and GCI (r= 0.4074), the others were all below 0.4. Overall, NDRE had the highest correlation with chlorophyll content with the two growth periods of potato plants. Interestingly, the correlation of NDRE in the V1 stage

 doi:10.20944/preprints202404.0311.v1

was higher than that of in the V2 stage. This is mainly due to the saturation phenomenon described in the introduction section.

When it considering texture features, mean value of blue band (B-mean) produced significant correlation with chlorophyll content (r= 0.8541). However, the variance of near-infrared band (RE-var) had best performing (r=-0.6829) at V2 stage. Moreover, there were no features observed in producing high correlation. As potato plants transition from tuber formation to tuber enlargement, the increase of potato flowers has produced more complex canopy texture features, which further lead to inconsistent performance of texture features between two growth periods.

Although at V1 stage, higher correlation was found in B-mean and R-mean than NDRE (Figure 3(a)), the above two parameters did not achieve ideal performance in V2 stage. Furthermore, at V2 stage, RE-var obtained higher correlation than NDRE. However, NDRE achieved the most significant correlation coefficient compared with other remaining parameters (Figure 3(b)). Taking into account the performance of all the above-mentioned features of two growth periods of potato plants, NDRE was selected to construct a new feature parameter.



(a) (b)

Figure 3. Feature screening results based on Pearson correlation coefficient method. (a) Importance ranking of V1 stage, (b) Importance ranking of V2 stage.

3.2. Feature Construction Results

PCA was conducted to screen the reflectivity of 5 bands, the first principal component analysis result (PCA1) and second principal component analysis result (PCA2) was positively correlated with NDRE respectively, which were not accessible for the combination of NDRE with PCA1 or PCA2. Figure 4(a) shows the third principal component analysis result (PCA3) have a negative correlation distribution law with NDRE. Figure 4(b) illustrates the NDRE and PCA3 have obvious linear relationship, with R^2 value of 0.9957 and RMSE value of 0.0045.

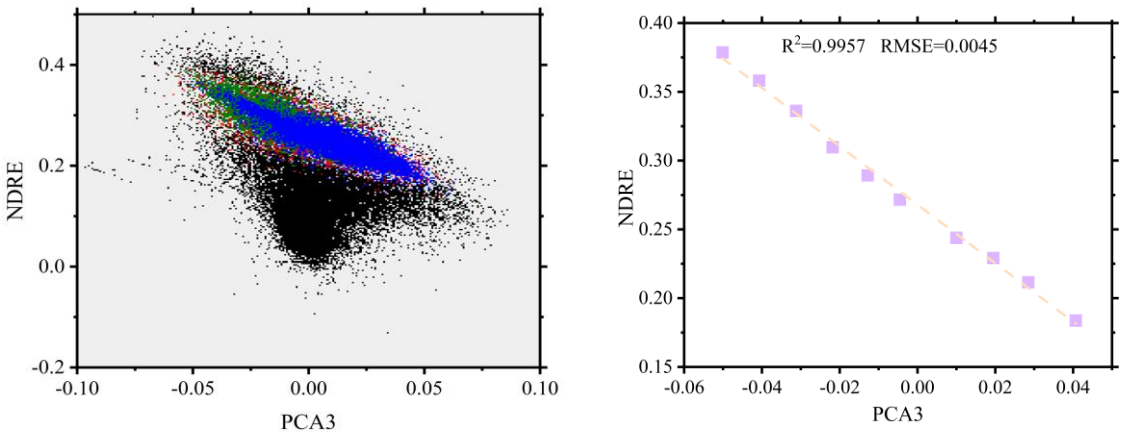
The chlorophyll content of potato plants in the experimental field increases with the increase of N application rate. This found was similar with [17,40]. In order to further describe the experimental field distribution of the above three PCAs, we generated their corresponding prescription maps. It could be observed from Figure 5(a) and Figure 5(b), PCA1 and PCA2 cannot be used to characterized distribution of chlorophyll content of potato plants. These are mainly influenced by different potato varieties. Figure 5(c), PCA3 per plot roughly decreased with N fertilizer application rate increasing, a response that was inverse to chlorophyll content. Therefore, combining NDRE with PCA3 can further improve the estimation accuracy. Improved NDRE (INDRE) was designed based on construction principle of NDRE, and the calculation formula for INDRE was shown as Eq. (1).

$$INDRE = \frac{PCA3 - NDRE}{PCA3 + NDRE} \quad (1)$$

At V1 and V2 stage, the correlation coefficient of INDER was 0.8102 and 0.7792 respectively. Compared with NDRE, the r value increased by 0.2091 and 0.2469 respectively. Moreover, the correlation coefficient of INDRE was only lower than B-mean at V1 stage, and at V2 stage, INDRE achieved the highest correlation coefficient. It can be seen that the INDRE proposed by combining

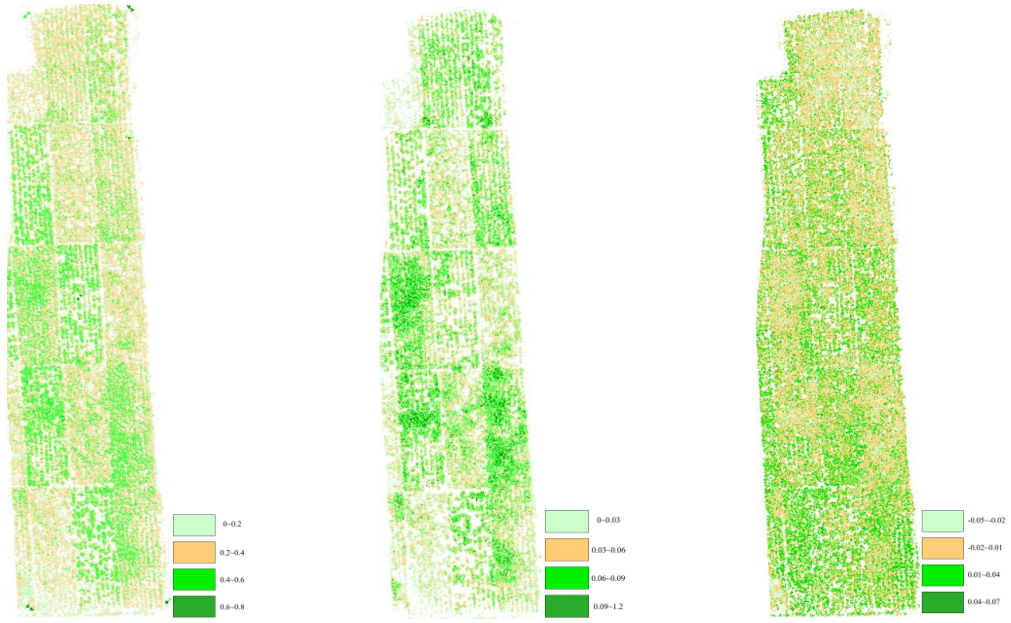
Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 April 2024 doi:10.20944/preprints202404.0311.v1

the PCA3 of the above 5 bands with the NDRE construction principle has great potential to improve the correlation.



(a) (b)

Figure 4. (a) Scatter distribution of potato plants under NDRE and PCA3 fitting conditions, (b) Linear fitting results of potato plants in NDRE and PCA3.



(a) (b) (c)

Figure 5. Distribution results of the first three PCAs in the experimental field for (a) PCA1, (b) PCA2 and (c) PCA3.

3.3. Results of Chlorophyll Content Estimation

During the two growth periods, INDRE, NDRE, B-mean, and RE-var with high correlation were used to establish inversion model of chlorophyll content, respectively. Potato plants are less affected by the growth period at a single stage, so linear regression models are applied to estimate chlorophyll content. On the contrary, as the potato plant grows, its physical and chemical indicators will produce significant changes, further leading to a huge difference in the canopy spectrum. Therefore, the polynomial regression models were selected to estimate chlorophyll content for two growth periods. At V1 stage, B-mean obtained the highest estimation accuracy of chlorophyll content (R^2=0.7296, RMSE=2.3409) followed by INDRE (R^2=0.6565, RMSE=2.6382), NDRE (R^2=0.3613, RMSE=3.5973) and RE-var (R^2=0.1353, RMSE=4.1858).

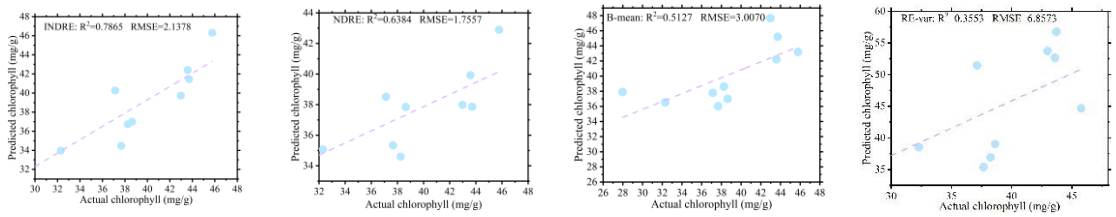
 doi:10.20944/preprints202404.0311.v1

At V2 stage, INDRE produced the best performance (R^2=0.6072, RMSE=2.3453), followed by RE-var (R^2=0.4663, RMSE=2.7337), NDRE (R^2=0.2833, RMSE=3.1679) and B-mean (R^2=0.1377, RMSE=3.4747). Although the chlorophyll content estimation accuracy of INDRE at V1 stage was lower than that of B-mean, the best accuracy was achieved at V2 stage. To further evaluate the performance of the above four parameters capacity, chlorophyll content measured in the two growth periods was used to assess the model. Figure.6 demonstrates INDRE yield the best chlorophyll content estimation accuracy (R^2=0.7865, RMSE=2.5132), followed by NDRE (R^2=0.6384, RMSE=3.8126), B-mean (R^2=0.5127, RMSE=3.9386) and RE-var (R^2=0.3553, RMSE=8.6215), and the corresponding R^2 and RMSE increases by 0.1481 and decreases by 1.2994 respectively compared with NDRE. When the chlorophyll content measured at different growth periods of potato plants were applied, the variation of which increased, so INDVI obtained a higher estimation accuracy compared with a single stage. Moreover, INDRE yielded higher estimation accuracy compared to other variables, which further proved the feasibility of combining PCA.

4. Discussion

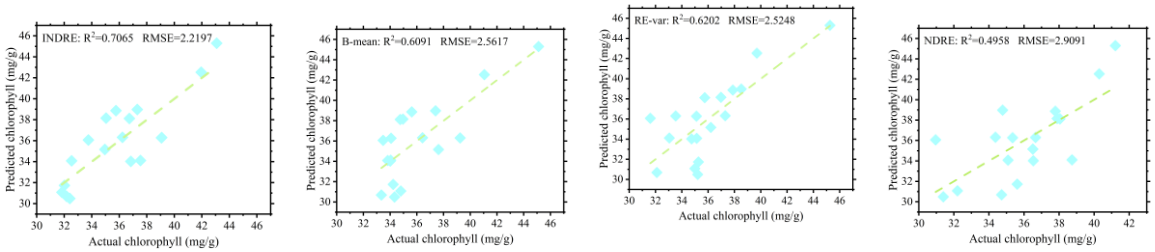
4.1. Universality of Chlorophyll Content Estimation Models

When the observation environment changes, the accuracy of chlorophyll content estimation will be greatly affected [19]. Therefore, it is imperative to evaluate the chlorophyll inversion models of potato plants measured in different environments to verify their universality. Generally, there were lower accuracy of regression model of chlorophyll content established in other environments than that of modeling subset due to the time, locations, and scales influence. Soil type, planting method and growth environment are different among Exp. 1 and Exp. 2. In order to assess universality, we applied above parameters with strong correlation to plant chlorophyll content in Exp.1 as standalone data obtained in Exp. 2. Figure 7 were the comparisons between measured and predicted chlorophyll content of the INDRE, NDRE, B-mean and RE-var. Figure 6 indicates that INDRE similarly obtain the highest estimation accuracy of chlorophyll content with R^2 value of 0.7065 and RMSE value of 2.2197, followed by RE-var (R^2=0.6202, RMSE=2.5248), B-mean (R^2=0.6091, RMSE=2.5617) and NDRE (R^2=0.4958, RMSE=2.9091). The proposed INDRE performed well in the estimation of potato plants chlorophyll content in different growth environments.



(a) (b) (c) (d)

Figure 6. Validation of estimation model based on different parameters in Exp. 1.



(**a**) (**b**) (c) (d)

Figure 7. Validation of estimation model based on different parameters in Exp.2, for (a) NDRE, (b) INDRE, (c) B-mean and (d) RE-mean.

 doi:10.20944/preprints202404.0311.v1

4.2. Comparison of Different Feature Selection Algorithms

The Pearson correlation coefficient method was used to screen variables by evaluating the linear relationship between two parameters. Apart from Pearson correlation coefficient method, random forest (RF) provides feasible solutions for feature screening by comparing feature contributions [18,41]. RF was selected to assess the contribution rate of each feature during the two growth periods of potato plants in comparison with Pearson correlation coefficient method to further explore the law of feature contribution. For the convenience of readers to observe the contribution of each parameter more vividly, the first 8 features with higher contribution were displayed. Although, Figure 8 shows that the screening results of RF are inconsistent with those of Pearson correlation coefficient method, based on the results of RF screening in two growth periods, there were closer relationship between NDRE and chlorophyll content. Therefore, it is reasonable to use NDRE for subsequent feature construction.

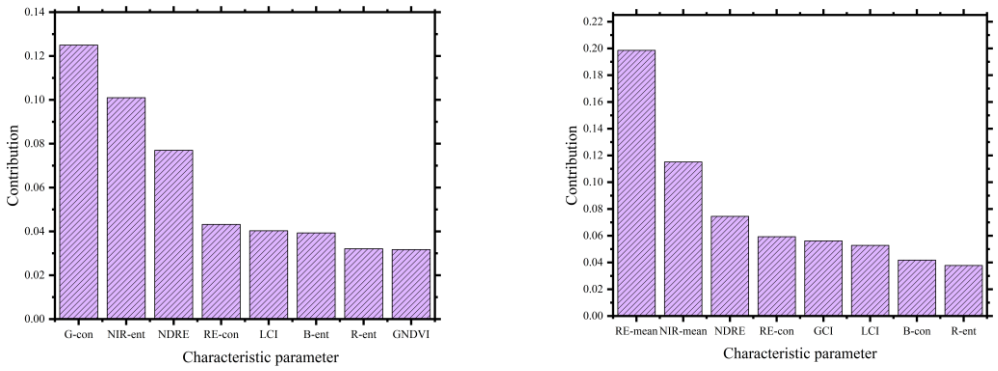


Figure 8. Feature screening results based on RF algorithm. (a) Importance ranking of V1 stage;(b) Importance ranking of V2 stage.

4.3. The Effect of Fitting Parameter Selection on Estimation Accuracy

Different fitting parameters had a great influence on the results of INDRE construction, so an inappropriate setup of the fitting manner generally lead to misleading estimated results [42]. K-means clustering was conduct to classify experimental fields, and number of clusters was set to be 5 according the gradient of N fertilizer application rate with 20 iterations, the corresponding five categories of clustering results were named C 1, C 2, C 3, C 4 and C 5. Figure 9 (b) and Figure 9(c) indicates separately that the clustering results of PCA1 and PCA2 sampling plots is greatly influenced by potato varieties. Unlike PCA1 and PCA2, the clustering results of PCA3 are mainly affected by the amount of N fertilizer applied instead of potato varieties (Figure 9 (d)). To provide a more detailed description of this difference, the proportion of clustering results of different PCAs were compared by least significant different test (LSD) at the 95% level of significance. Figure 10, there were no significant difference in PCA1 of C 3, C 4 and C 5. In addition, there were significant differences among the three potato varieties in other categories. For PCA2, there were significant difference among the three varieties of C 1, C 3, and C 4, and the remaining other categories had no significant difference. In PCA3, C 1 and C 3 also had significant difference of three potato varieties, this may be influenced by the amount of N fertilizer applied. Interestingly, significant differences in the proportion of C 1 among the PCAs between potato varieties were found.

Similarly, the significant differences among the 5 categories under the 5 N application gradients were analyzed. Figure 11 shows no significant differences are found in the 5 categories of PCA1 and PCA2. On the contrary, there were significant difference among the 5 N fertilizer gradients of C 1, C 2, and C 3 in the PCA3. The PCA3 of C 1 was roughly negatively correlated with N application rate and there was no obvious pattern in C2 and C3. At the same time, we analyzed the reasons for the differences between three potato varieties in PCA3. Figure 12 indicates that the differences between potato varieties are basically consistent with the changes in N fertilization application gradient in C

 doi:10.20944/preprints202404.0311.v1

1 and C 3. Therefore, it can be inferred that the differences between potato varieties are caused by the amount of N fertilizer applied.

To fully excavated the UAV- based multispectral images and evaluate the applicability of PCA1 and PCA2, the fitting results of the PCA1 and the PCA2 with NDRE were used to establish prediction model. Figure 13 shows both PCA1 and PCA2 have a significant linear distribution with NDRE, and the corresponding R^2 exceeds 0.99. Regretly, the both INDRE constructed separately based on PCA1 and PCA2 have unsatisfactory performance (Figure 14). However, INDRE based on PCA3 could well characterize the chlorophyll content of potato plants, and achieved better estimation accuracy compared with other common parameters.

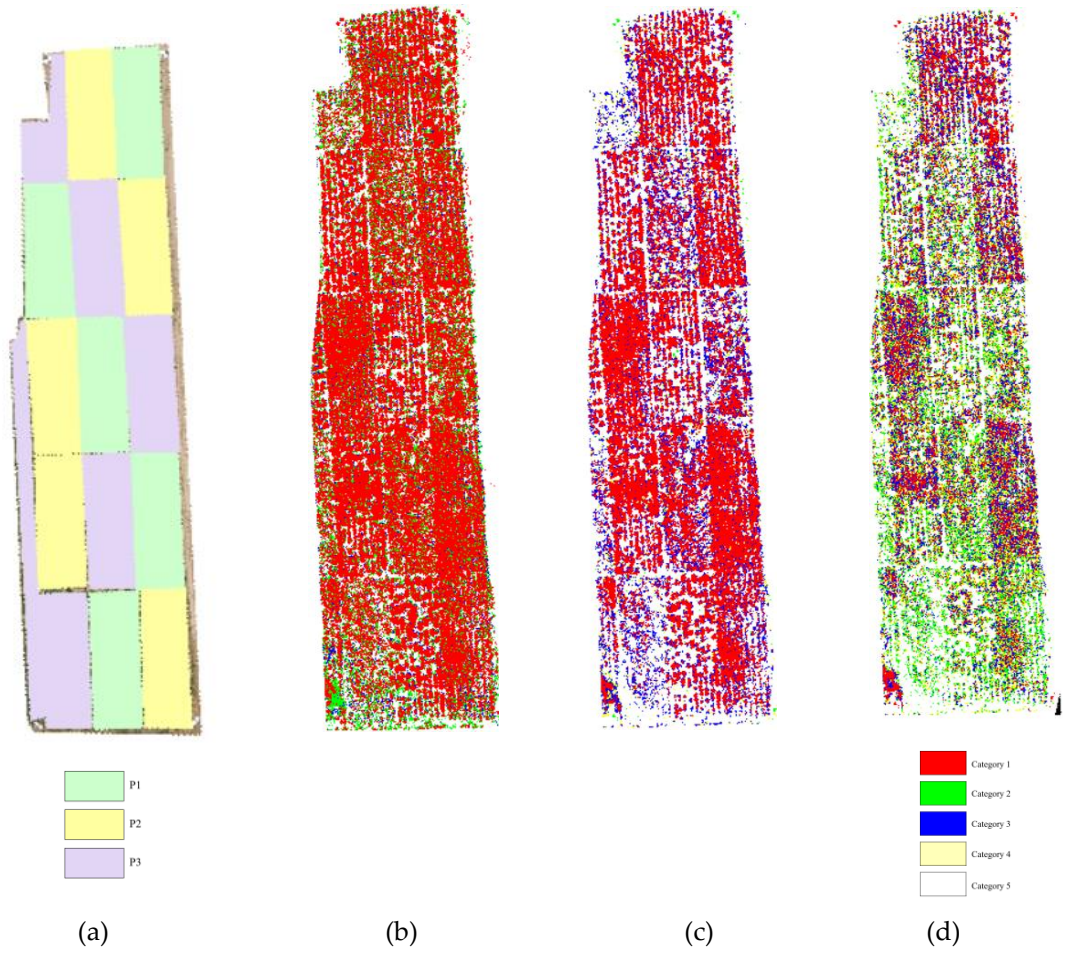
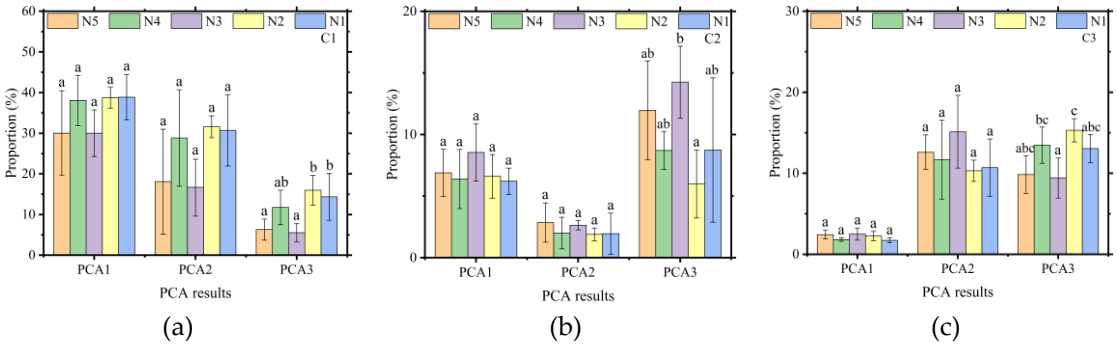
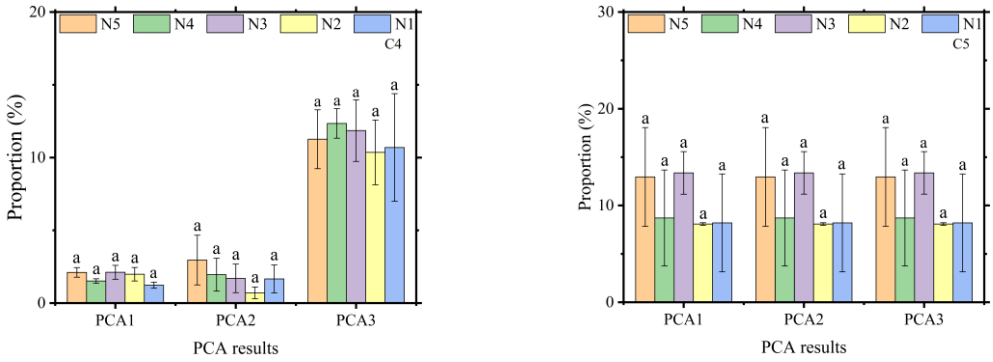


Figure 9. Classification results of experimental field based on K-means cluster for (a) distribution of three potato varieties, (b) PCA1, (c) PCA2, (d) PCA3.

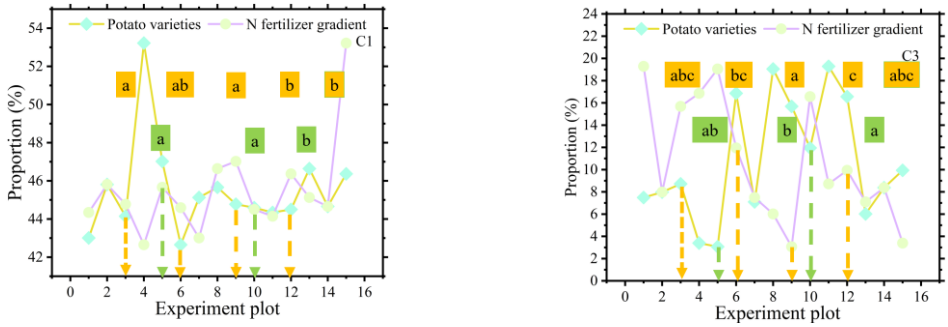


Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 3 April 2024 doi:10.20944/preprints202404.0311.v1



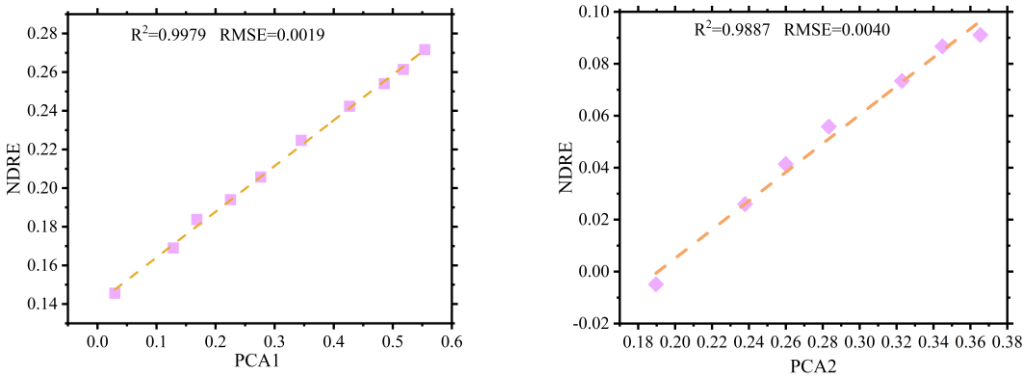
(d) (e)

Figure 11. Significance analysis of the proportion of five cluster categories under the first three principal component analysis results of five N fertilizer gradients. (a) C 1, (b) C 2, (c) C 3, (d) C 4 and (e) C 5. The corresponding significant differences had been represented by different letters ($p < 0.05$).



(a) (b)

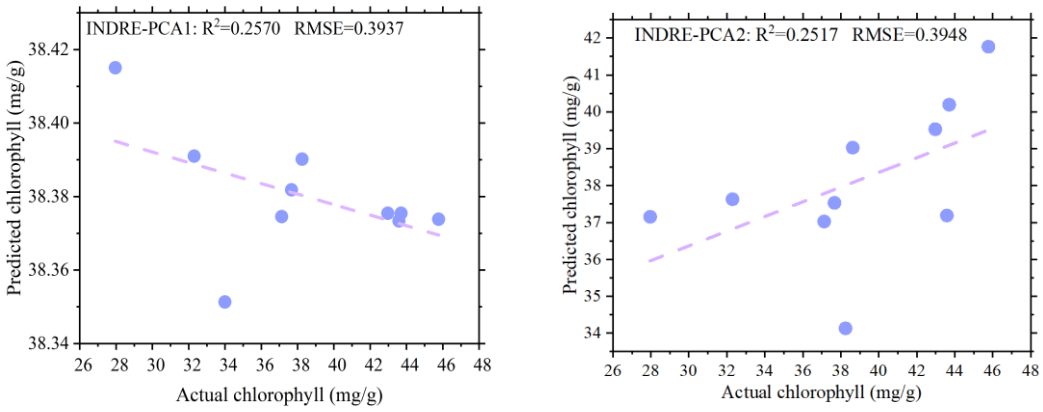
Figure 12. Comparison results of significant differences between C 1 and C 3 for potato varieties and N fertilizer gradients, for (a) C 1 and (b) C 3.



(a) (b)

Figure 13. Relationship between different PCA results and NDRE. (a) The linear relationship between PCA1 and NDRE;(b) The linear relationship between PCA2 and NDRE.

 doi:10.20944/preprints202404.0311.v1



(a) (b)

Figure 14. Validation of chlorophyll content estimation model with INDRE constructed based on PCA1 and PCA2 respectively, for (a) INDRE-PCA1 and (b) INDRE-PCA2.

4.4. Significance, Advantages, and Disadvantages of This Study

Under similar control conditions, Qiao et al. used wavelet analysis to remove soil background of UAV-based multispectral images, and then established a regression model for maize canopy chlorophyll content estimation based on PLS method [33]. The results showed that removing soil background could improve the accuracy of chlorophyll content estimation with the highest R^2 of 0.6923. Yang et al. extracted common multispectral VIs, texture features and fractional vegetation cover (FVC) of potato plants, and constructed an estimation model of chlorophyll content based on stacking algorithms (R^2 = 0.739, RMSE =0.511)[20]. Although various machine learning algorithms had been fully excavated, VIs and variables combination manner, the multispectral images information does not fully retain, resulting in a lower accuracy of chlorophyll content estimation. Therefore, it is necessary to improve the accuracy of estimating chlorophyll content by fully excavate the UAV-based multispectral images information.

This study combining PCA results of five multispectral bands with NDRE performed well with highest R^2 value of 0.7865, RMSE value of 2.1378. Moreover, measured chlorophyll content in Exp. 2 was used to evaluate the accuracy and robustness of this model, unfortunately, the model accuracy decreased, but it was still in an acceptable range in Exp. 2. The difference in soil, management methods, and climatic conditions between Exp. 1 and Exp. 2 mainly caused this phenomenon. In subsequent research, we will collect chlorophyll content data from multiple factors to further improve the accuracy and universality of potato chlorophyll content inversion models.and PCA2 respectively, for (a) INDRE-PCA1 and (b) INDRE-PCA2.

5. Conclusions

To fully utilize the UAV-based multispectral images information to estimating the chlorophyll content of potato plants, the INDRE constructed based on the combining NDER with the 3rd PCA results for the 5 above-mentioned bands was proposed in this paper. INDRE was fully analyzed and validated against measured chlorophyll content, and compared with common parameters, including VIs and texture features, with main conclusions as following:

(1) The screening results of the Pearson correlation coefficient method showed that the NDRE had the highest correlation with chlorophyll content at 75th day after potato planting. Although the feature screening results of RF showed that the contribution of NDRE in the two growth periods of potato plants was not the highest, combining the screening results of the two periods, NDRE had achieved well performance.

(2) The PCA1 and PCA 2 of experimental field were greatly influenced by potato varieties, and the PCA3 was negatively correlated with N application rate. Moreover, NDRE had a negative correlation with the PCA3, which could be effectively combined.

 doi:10.20944/preprints202404.0311.v1

(3) The INDRE, proposed on the basis of the the construction principle of NDRE, and combining the NDRE with the PCA3 significantly improved the estimation accuracy of chlorophyll content. However, the INDRE constructed separately based on the PCA1 and PCA2 did not achieve promising performance. Besides, the model used for chlorophyll content retrieval in Exp.1 also performed well in Exp.2, which proved that the INDRE proposed in this paper had better estimation accuracy and robustness of chlorophyll content in potato plants.

Author Contributions: Conceptualization, H.Y. and Y.H.; methodology, H.Y.; formal analysis, H.Y., X.S. and K.Z.; data curation, H.Y.; writing—original draft H.Y., X.S and T.G.; writing—review and editing, H.Y. and X.S.; visualization, H.Y. and T.G.; supervision, H.Y. and Y.H.; project administration, H.Y. and Y.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China (32171894 (C 0043619), 31971787(C 0043628)).

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful to Yichen Qiao, Penghui Liu, Manshuai Zhang and Peng Zhang for data collection.

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

References

1. An, L.; Tang, W.; Qiao, L.; Zhao, R.; Sun, H.; Li, M.; Zhang, Y.; Zhang, M.; Li, X., Estimation of chlorophyll distribution in banana canopy based on RGB-NIR image correction for uneven illumination. Computers and Electronics in Agriculture 2022, 202, 107358.
2. Zhao, R.; An, L.; Tang, W.; Gao, D.; Qiao, L.; Li, M.; Sun, H.; Qiao, J., Deep learning assisted continuous wavelet transform-based spectrogram for the detection of chlorophyll content in potato leaves. Computers and Electronics in Agriculture 2022, 195, 106802.
3. Jia, M.; Colombo, R.; Rossini, M.; Celesti, M.; Zhu, J.; Cogliati, S.; Cheng, T.; Tian, Y.; Zhu, Y.; Cao, W.; Yao, X., Estimation of leaf nitrogen content and photosynthetic nitrogen use efficiency in wheat using sun-induced chlorophyll fluorescence at the leaf and canopy scales. European Journal of Agronomy 2021, 122, 126192.
4. Du, P.; Yin, B.; Zhou, S.; Li, Z.; Zhang, X.; Cao, Y.; Han, R.; Shi, C.; Liang, B.; Xu, J., Melatonin and dopamine mediate the regulation of nitrogen uptake and metabolism at low ammonium levels in Malus hupehensis. Plant Physiology and Biochemistry 2022, 171, 182-190.
5. Wang, N.; Fu, F.; Wang, H.; Wang, P.; He, S.; Shao, H.; Ni, Z.; Zhang, X., Effects of irrigation and nitrogen on chlorophyll content, dry matter and nitrogen accumulation in sugar beet (Beta vulgaris L.). Scientific Reports 2021, 11, (1), 16651.
6. Qiao, L.; Tang, W.; Gao, D.; Zhao, R.; An, L.; Li, M.; Sun, H.; Song, D., UAV-based chlorophyll content estimation by evaluating vegetation index responses under different crop coverages. Computers and Electronics in Agriculture 2022, 196, 106775.
7. Jiang, X.; Zhen, J.; Miao, J.; Zhao, D.; Shen, Z.; Jiang, J.; Gao, C.; Wu, G.; Wang, J., Newly-developed three-band hyperspectral vegetation index for estimating leaf relative chlorophyll content of mangrove under different severities of pest and disease. Ecological Indicators 2022, 140, 108978.
8. Zhu, C.; Ding, J.; Zhang, Z.; Wang, J.; Wang, Z.; Chen, X.; Wang, J., SPAD monitoring of saline vegetation based on Gaussian mixture model and UAV hyperspectral image feature classification. Computers and Electronics in Agriculture 2022, 200, 107236.
9. Meiyan, S.; Mengyuan, S.; Qizhou, D.; Xiaohong, Y.; Baoguo, L.; Yuntao, M., Estimating the maize above-ground biomass by constructing the tridimensional concept model based on UAV-based digital and multi-spectral images. Field Crops Research 2022, 282, 108491.
10. Jin, X.; Zarco-Tejada, P. J.; Schmidhalter, U.; Reynolds, M. P.; Hawkesford, M. J.; Varshney, R. K.; Yang, T.; Nie, C.; Li, Z.; Ming, B.; Xiao, Y.; Xie, Y.; Li, S., High-Throughput Estimation of Crop Traits: A Review of Ground and Aerial Phenotyping Platforms. IEEE Geoscience and Remote Sensing Magazine 2021, 9, (1), 200-231.

 doi:10.20944/preprints202404.0311.v1

11. Yang, X.; Yang, R.; Ye, Y.; Yuan, Z.; Wang, D.; Hua, K., Winter wheat SPAD estimation from UAV hyperspectral data using cluster-regression methods. International Journal of Applied Earth Observation and Geoinformation 2021, 105, 102618.
12. Fu, Y.; Yang, G.; Song, X.; Li, Z.; Xu, X.; Feng, H.; Zhao, C., Improved estimation of winter wheat aboveground biomass using multiscale textures extracted from UAV-based digital images and hyperspectral feature analysis. Remote Sensing 2021, 13, (4), 581.
13. Wang, Z.; Chen, J.; Zhang, J.; Tan, X.; Raza, M. A.; Ma, J.; Zhu, Y.; Yang, F.; Yang, W., Assessing canopy nitrogen and carbon content in maize by canopy spectral reflectance and uninformative variable elimination. The Crop Journal 2022, 10, (5), 1224-1238.
14. Guo, Y.; Chen, S.; Li, X.; Cunha, M.; Jayavelu, S.; Cammarano, D.; Fu, Y., Machine Learning-Based Approaches for Predicting SPAD Values of Maize Using Multi-Spectral Images. In Remote Sensing, 2022; Vol. 14.
15. Sudu, B.; Rong, G.; Guga, S.; Li, K.; Zhi, F.; Guo, Y.; Zhang, J.; Bao, Y., Retrieving SPAD values of summer maize using UAV hyperspectral data based on multiple machine learning algorithm. Remote Sensing 2022, 14, (21), 5407.
16. J.; P.; G.; Rigon; S.; Capuani; D.; M.; Fernandes; T., A novel method for the estimation of soybean chlorophyll content using a smartphone and image analysis. Photosynthetica 2016, 54, (4), 559–566.
17. Yang, H.; Li, J.; Yang, J.; Hua, W.; Zou, J.; He, J.; Hui, D., Effects of Nitrogen Application Rate and Leaf Age on the Distribution Pattern of Leaf SPAD Readings in the Rice Canopy. Plos One 2014, 9, (2), e88421.
18. Yang, H.; Hu, Y.; Zheng, Z.; Qiao, Y.; Hou, B.; Chen, J., A New Approach for Nitrogen Status Monitoring in Potato Plants by Combining RGB Images and SPAD Measurements. Remote Sensing 2022, 14, (19), 4814.
19. Qi, H.; Wu, Z.; Zhang, L.; Li, J.; Zhou, J.; Jun, Z.; Zhu, B., Monitoring of peanut leaves chlorophyll content based on drone-based multispectral image feature extraction. Computers and Electronics in Agriculture 2021, 187, 106292.
20. Yang, H.; Hu, Y.; Zheng, Z.; Qiao, Y.; Zhang, K.; Guo, T.; Chen, J., Estimation of Potato Chlorophyll Content from UAV Multispectral Images with Stacking Ensemble Algorithm. Agronomy 2022, 12, (10), 2318.
21. Singhal, G.; Bansod, B.; Mathew, L.; Goswami, J.; Choudhury, B. U.; Raju, P. L. N., Chlorophyll estimation using multi-spectral unmanned aerial system based on machine learning techniques. Remote Sensing Applications: Society and Environment 2019, 15, 100235.
22. Zhang, Y.; Hui, J.; Qin, Q.; Sun, Y.; Zhang, T.; Sun, H.; Li, M., Transfer-learning-based approach for leaf chlorophyll content estimation of winter wheat from hyperspectral data. Remote Sensing of Environment 2021, 267, 112724.
23. Sun, Q.; Gu, X.; Chen, L.; Xu, X.; Wei, Z.; Pan, Y.; Gao, Y., Monitoring maize canopy chlorophyll density under lodging stress based on UAV hyperspectral imagery. Computers and Electronics in Agriculture 2022, 193, 106671.
24. Shu, M.; Shen, M.; Zuo, J.; Yin, P.; Wang, M.; Xie, Z.; Tang, J.; Wang, R.; Li, B.; Yang, X.; Ma, Y., The Application of UAV-Based Hyperspectral Imaging to Estimate Crop Traits in Maize Inbred Lines. Plant Phenomics 2021.
25. Shao, G.; Han, W.; Zhang, H.; Wang, Y.; Zhang, L.; Niu, Y.; Zhang, Y.; Cao, P., Estimation of transpiration coefficient and aboveground biomass in maize using time-series UAV multispectral imagery. The Crop Journal 2022, 10, (5), 1376-1385.
26. Yang, H.; Lan, Y.; Lu, L.; Gong, D.; Miao, J.; Zhao, J., New method for cotton fractional vegetation cover extraction based on UAV RGB images. International Journal of Agricultural and Biological Engineering 2022, 15, (4), 172-180.
27. Hashemi-Nasab, F. S.; Parastar, H., Vis-NIR hyperspectral imaging coupled with independent component analysis for saffron authentication. Food Chemistry 2022, 393, 133450.
28. Shao, Y.; Shi, Y.; Xuan, G.; Li, Q.; Wang, F.; Shi, C.; Hu, Z., Hyperspectral imaging for non-destructive detection of honey adulteration. Vibrational Spectroscopy 2022, 118, 103340.
29. González-Jaramillo, V.; Fries, A.; Bendix, J., AGB Estimation in a Tropical Mountain Forest (TMF) by Means of RGB and Multispectral Images Using an Unmanned Aerial Vehicle (UAV). Remote Sensing 2019, 11, (12), 1413-.
30. Tavakoli, H.; Gebbers, R., Assessing Nitrogen and water status of winter wheat using a digital camera. Computers and Electronics in Agriculture 2019, 157, 558-567.

 doi:10.20944/preprints202404.0311.v1

31. Kaufman, Y. J.; Tanre, D., Atmospherically resistant vegetation index (ARVI) for EOS-MODIS. IEEE Transactions on Geoscience and Remote Sensing 1992, 30, (2), 261-270.
32. Gilabert, M. A.; González-Piqueras, J.; García-Haro, F. J.; Meliá, J., A generalized soil-adjusted vegetation index. Remote Sensing of Environment 2002, 82, (2), 303-310.
33. Qiao, L.; Gao, D.; Zhang, J.; Li, M.; Sun, H.; Ma, J., Dynamic influence elimination and chlorophyll content diagnosis of maize using UAV spectral imagery. Remote Sensing 2020, 12, (16), 2650.
34. Yan, K.; Gao, S.; Chi, H.; Qi, J.; Song, W.; Tong, Y.; Mu, X.; Yan, G., Evaluation of the Vegetation-Index-Based Dimidiate Pixel Model for Fractional Vegetation Cover Estimation. IEEE Transactions on Geoscience and Remote Sensing 2022, 60, 1-14.
35. Owen, T. W.; Carlson, T. N.; Gillies, R. R., An assessment of satellite remotely-sensed land cover parameters in quantitatively describing the climatic effect of urbanization. International Journal of Remote Sensing 1998, 19, (9), 1663-1681.
36. García Cárdenas, D. A.; Ramón Valencia, J. A.; Alzate Velásquez, D. F.; Palacios Gonzalez, J. R. In Dynamics of the Indices NDVI and GNDVI in a Rice Growing in Its Reproduction Phase from Multi-spectral Aerial Images Taken by Drones, Advances in Information and Communication Technologies for Adapting Agriculture to Climate Change II, Cham, 2019//, 2019; Corrales, J. C.; Angelov, P.; Iglesias, J. A., Eds. Springer International Publishing: Cham, 2019; pp 106-119.
37. Virnodkar, S. S.; Pachghare, V. K.; Patil, V. C.; Jha, S. K., Remote sensing and machine learning for crop water stress determination in various crops: a critical review. Precision Agriculture 2020, 21, (5), 1121-1155.
38. Mahamid, F. A.; Bdier, D., The Association Between Positive Religious Coping, Perceived Stress, and Depressive Symptoms During the Spread of Coronavirus (COVID-19) Among a Sample of Adults in Palestine: Across Sectional Study. Journal of Religion and Health 2021, 60, (1), 34-49.
39. Wang, T.; Gao, M.; Cao, C.; You, J.; Zhang, X.; Shen, L., Winter wheat chlorophyll content retrieval based on machine learning using in situ hyperspectral data. Computers and Electronics in Agriculture 2022, 193, 106728.
40. Li, R.; Chen, J.; Qin, Y.; Fan, M., Possibility of using a SPAD chlorophyll meter to establish a normalized threshold index of nitrogen status in different potato cultivars. Journal of Plant Nutrition 2019, 1-8.
41. Gomes, L. C.; Faria, R. M.; De Souza, E.; Veloso, G. V.; Schaefer, C. E. G. R.; Filho, E. I. F., Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 2019, 340, 337-350.
42. Zhou, Z.; Plauborg, F.; Thomsen, A. G.; Andersen, M. N., A RVI/LAI-reference curve to detect N stress and guide N fertigation using combined information from spectral reflectance and leaf area measurements in potato. European Journal of Agronomy 2017, 87, 1-7.