



Article

Assessment of Regression Models for Predicting Rice Yield and Protein Content Using Unmanned Aerial Vehicle-Based Multispectral Imagery

Yeseong Kang ¹, Jinwoo Nam ², Younggwang Kim ², Seongtae Lee ², Deokgyeong Seong ², Sihyeong Jang ¹ and Chanseok Ryu ^{1,*}

¹ Department of Bio-System Engineering, Gyeongsang National University, Jinju-si 52828, Korea; slow321@gnu.ac.kr (Y.K.); jjangjmh@nate.com (S.J.)

² Gyeongnam Agricultural Research & Extension Services, Jinju-si 52733, Korea; fireant84@korea.kr (J.N.); kimykw@korea.kr (Y.K.); lst08@korea.kr (S.L.); sdk201@korea.kr (D.S.)

* Correspondence: ryucs@gnu.ac.kr

Abstract: Unmanned aerial vehicle-based multispectral imagery including five spectral bands (blue, green, red, red-edge, and near-infrared) for a rice field in the ripening stage was used to develop regression models for predicting the rice yield and protein content and to select the most suitable regression analysis method for the year-invariant model: partial least squares regression, ridge regression, and artificial neural network (ANN). The regression models developed with six vegetation indices (green normalization difference vegetation index (GNDVI), normalization difference red-edge index (NDRE), chlorophyll index red edge (CIrededge), difference NIR/Green green difference vegetation index (GDVI), green-red NDVI (GRNDVI), and medium resolution imaging spectrometer terrestrial chlorophyll index (MTCI)), calculated from the spectral bands, were applied to single years (2018, 2019, and 2020) and multiple years (2018 + 2019, 2018 + 2020, 2019 + 2020, and all years). The regression models were cross-validated through mutual prediction against the vegetation indices in nonoverlapping years, and the prediction errors were evaluated via root mean squared error of prediction (RMSEP). The ANN model was reproducible, with low and sustained prediction errors of $24.2 \text{ kg}/1000 \text{ m}^2 \leq \text{RMSEP} \leq 59.1 \text{ kg}/1000 \text{ m}^2$ in rice yield and $0.14\% \leq \text{RMSEP} \leq 0.28\%$ in rice-protein content in all single-year and multiple-year analyses. When the importance of each vegetation index of the regression models was evaluated, only the ANN model showed the same ranking in the vegetation index of the first (MTCI in both rice yield and protein content) and second importance (CIrededge in rice yield and GRNDVI in rice-protein content). Overall, this means that the ANN model has the highest potential for developing a year-invariant model with stable RMSEP and consistent variable ranking.

Keywords: multispectral imagery; mutual prediction; regression model; rice-protein content; rice yield



Citation: Kang, Y.; Nam, J.; Kim, Y.; Lee, S.; Seong, D.; Jang, S.; Ryu, C. Assessment of Regression Models for Predicting Rice Yield and Protein Content Using Unmanned Aerial Vehicle-Based Multispectral Imagery. *Remote Sens.* **2021**, *13*, 1508. <https://doi.org/10.3390/rs13081508>

Academic Editor: Xia Yao

Received: 1 March 2021

Accepted: 13 April 2021

Published: 14 April 2021

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



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

1. Introduction

Crop modeling, which is a method of obtaining quantitative knowledge of how crops grow by interacting with the environment, is a useful tool for predicting the yield and quality to help with crop management [1,2]. Developed crop models include climatic variables such as the temperature, precipitation, and length of day [3]. However, crop models are only applicable to normal patterns of the given climatic conditions, and all other conditions are assumed optimal [1]. Therefore, crop models cannot provide realistic predictions of the crop yield or quality. Remote sensing (RS) technology has been combined with spectroscopy for nondestructive monitoring of the crop yield and quality under actual environmental conditions [4]. Spectral reflectance data containing visible, red-edge, and near-infrared (NIR) spectral bands have tried to managing agriculture over a wide area [5]

by helping predict the growth status [6]. This technology is expected to play a key role in integrated crop management systems [7].

Rice (*Oryza sativa* L.) is a major carbohydrate source for meeting energy and nutrient needs, and consumption is steadily increasing globally [8]. It is an essential crop, especially in Asia, and its successful annual production has significant implications for the global community [9]. Properties such as the color, flavor, and composition of the rice depend on the variety, storage conditions, and amylose content. Interactions among proteins, the breakdown products of lipid oxidation, and starch change the cell walls and proteins, which affect the rice quality [10]. The rice quality is improved by reducing protein content and amylose [11], which form by nitrogen uptake during the vegetative growth period, especially in the grain filling stage.

To recognize a specific rice field, a spatially precise RS system is required rather than a conventional spectrometer [12]. Spectral imagery combines spectrometry and imaging, and it has been used to observe specific spatial variations [13]. Spectral imagery can be used to assess the crop growth in individual areas while excluding external disturbances [14]. It can be used to develop a crop model that includes various parameters, such as the biomass, crop quality, and physiological active substances of the crop, with a comprehensive range of crop–canopy attributes [15].

Regression models have been developed to predict crop parameters during the cultivation period [16]. Simple, multiple, and multivariate linear regression analyses have been used with RS spectral data for agricultural applications. Simple linear regression (SLR) has been used with satellite spectral imagery to predict rice yields where the yield is the dependent variable and the vegetation index (VI) is the independent variable [17]. Multiple linear regression (MLR) and partial least squares regression (PLSR) have been used with airborne and satellite spectral imagery to predict the protein and nitrogen contents of rice [18,19]. Unmanned aerial vehicles (UAVs) are suitable for precise monitoring of crop parameters with a high spatial resolution by flying at low altitudes, and they are advantageous for acquiring time series data with unconstrained scheduling [20]. The multispectral imagery acquired by UAVs has been used for developing linear regression models by combining VIs for predicting the rice yield or biomass depending on time series [21]. Various approaches have been attempted to predict rice yield by labeling pixels in the ear region using K-means clustering on RGB imagery as well as spectral imagery [22]. In the prediction of rice-protein content, which requires more spectroscopy than morphology, some studies have suggested that green normalization difference VI (GNDVI) calculated from NIR and green is the most advantageous [23,24].

Recently, machine learning (ML) has been used to predict rice parameters with a high level of accuracy [25]. Among studies using ML with spectral imagery acquired by UAVs, linear ML-based ridge regression (RR) was used to predict the nitrogen content of rice in a single year [26]. Additionally, a nonlinear ML-based artificial neural network (ANN) was used to predict the moisture content of rice to reduce economic losses regarding the yield, quality, and drying cost at harvest time in a single year [27]. Among the various regression methods for the prediction of rice parameters, it is important to select a regression method that can be developed as a year-invariant model. The potential of developing a year-invariant model must be verified through cross-validation for different years via the collection of multiyear data [28]. However, although the model is based on multiyear data, changes in the environmental conditions at various fields may make it difficult to reproduce the model [18]. Therefore, cumulative climate data during growth periods in each year may be required as input variables besides the spectral imagery to realize a year-invariant model [29]. In addition, incorporating three-dimensional (3D) technology may have a positive effect on predicting rice growth grown under various environmental conditions. A study showed the possibility of predicting the crop yields depending on different space-times using a 3D-convolution neural network in RGB imagery acquired by UAV [30].

The objective of this study was to develop a regression model for predicting the rice yield and protein content and to select the most suitable regression analysis method for the year-invariant model. UAV-based multispectral imagery data for rice canopies at the ripening stage were collected in 2018, 2019, and 2020. Different types of regression models were developed for single years (2018, 2019, and 2020) and multiyears (2018 + 2019, 2018 + 2020, 2019 + 2020, and all years): PLSR (linear regression), RR (ML-based linear regression), and ANN (ML-based nonlinear regression). The regression models were evaluated by comparing the prediction performances for the rice yield and protein content. The reproducibility was verified via cross-validation of the regression models for single- and multi-year analyses in nonoverlapping years through mutual prediction. Finally, the regression model with the highest potential for the year-invariant model was selected.

2. Materials and Methods

2.1. Experimental Design

2.1.1. Field Site and History

The experiment was conducted in a rice field of the Agricultural Research and Extension Services in Jinju-si, Gyeongsangnam-do, Republic of Korea for a 3 year period (2018–2020), as shown in Figure 1. The rice-field area was 1550 m². The primary soil type of the rice field was sandy loam, and the rice cultivar was Yeonghojinmi. This is a high-quality cultivar artificially derived from crossing the high-quality Hitomebre with Junam, which has high disease resistance. Yeonghojinmi is a mid-late maturing ecotype that is mainly cultivated in the southern part of the Republic of Korea [31]. Rice seedlings were mechanically transplanted at a density of 30 × 14 cm. Table 1 presents the rice-cultivation schedule, which includes seedling management, water management, fertilizer prescription, and harvest time. The fertilizer prescription was divided into six levels of nitrogen (N) fertilization: 0, 5, 7, 9, 11, and 17 kg/1000 m². Additionally, each level received 4.5 kg/1000 m² of phosphorous (P) and 5.7 kg/1000 m² of potassium (K) for all years. N fertilization was applied three times at a ratio of 50%, 20%, and 30%, corresponding to the basal dressing before seedling transplantation, topdressing in the tillering stage, and topdressing in the panicle initiation stage. All P fertilization was applied to the basal dressing, and K fertilization was applied at a ratio of 70%, 0%, and 30%. The climatic conditions, such as the growing degree days (GDD) and accumulated sunlight hours (ASH), were collected from a weather station located in the Agricultural Research and Extension Services. The GDD and ASH were 1046 °C and 884 h, respectively, for 2018; 1118 °C and 873 h, respectively, for 2019; and 1108 °C and 790 h, respectively, for 2020. The multispectral images were acquired at midday on 4 October 2018; 1 October 2019; and 5 October 2020. Immediately after image acquisition, sampling was performed three times in 2018 and four times in 2019 and 2020, depending on each nitrogen treatment.

Table 1. Rice-cultivation schedule and climatic conditions each year.

	All Years
Soaking seeds	4 May
Pre-germination of seeds	6 May
Draining rice field	8 May
Sowing seeds	13 May
Irrigation of rice field	23 May
Basal dressing	31 May
Transplanting rice seedlings	6 June
Spraying herbicide	10 June
Topdressing in tillering stage	17 June
Topdressing in panicle initiation stage	30 July

Table 1. Cont.

	2018	2019	2020
Unmanned aerial vehicles (UAV)-based multispectral remote sensing (RS)	4 October	1 October	5 October
Harvest	18 October	17 October	17 October
Growing degree day (°C)	1046	1118	1108
Accumulated daylight hours (h)	884	873	790

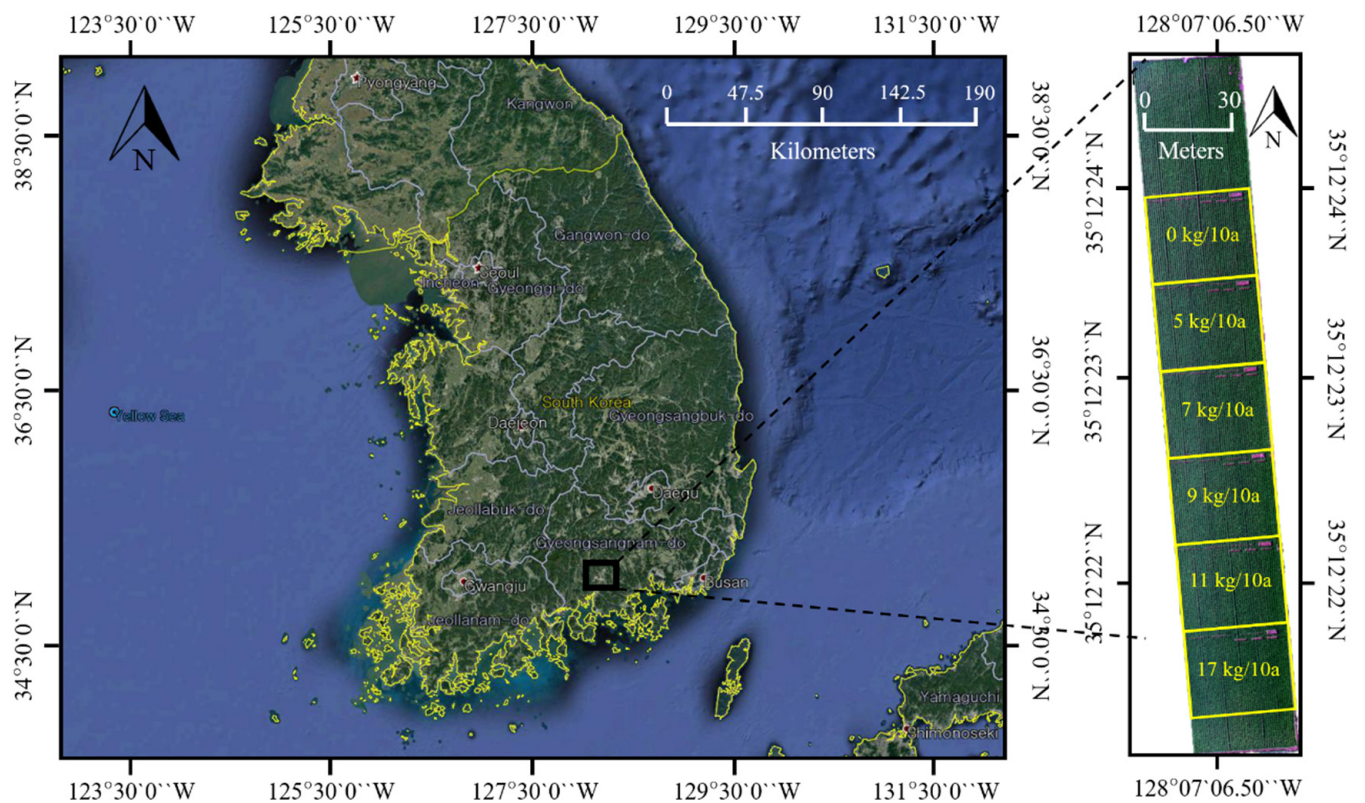


Figure 1. Location of rice field and layout of plots with different nitrogen fertilization treatments.

2.1.2. Measurement of the Rice Yield and Protein Content

To measure the rice yield and protein content, 100 and 15 rice plants were respectively sampled from each N fertilization treatment field just before harvest (18 October 2018; 17 October 2019; and 17 October 2020). The rice yield was measured by multiplying the number of rice ears per square meter (N), percent ripened grain (PRG), and unit seed weight (SW):

$$\text{Rice yield (kg/1000 m}^2\text{)} = N \times \text{PRG (\%)} \times \text{SW (g)} \times 1000 \text{ (m}^2\text{)} \div 1000 \quad (1)$$

The protein content of the polished rice was measured with a rice taste analyzer (Infratec 1241, FOSS Tecator, Hoganas, Sweden).

2.2. Acquisition of Multispectral Images

Multispectral images were acquired with a Red-edge-M (Micasense, Inc., Seattle, WA, USA) mounted on UAVs: a 3DR Solo (3D Robotics, Berkeley, CA, USA) in 2018 and a M600 (DJI Technology Co., Ltd., Shenzhen, China) in 2019 and 2020. The 3DR Solo and M600 are quadcopters weighing 1.5 and 9.1 kg, respectively, with dimensions of 0.26 m × 0.26 m × 0.25 m and 1.67 m × 1.52 m × 0.76 m, respectively. The flight plan software for the UAVs

were Mission Planner (ArduPilot Dev Team, New York, NY, USA) for the 3DR Solo and DJI Pilot (DJI Technology Co., Ltd., China) for the M600. Red-edge-M was equipped with a sunshine sensor to correct the reflectance data for each spectral image depending on the sunlight conditions at the time of image capture. The multispectral image and sunshine sensors were configured to measure the reflectance of the central wavelength \pm full width at half maximum for five spectral bands: blue (475 ± 32 nm), green (560 ± 27 nm), red (668 ± 14 nm), red-edge (717 ± 12 nm), and NIR (842 ± 57 nm). The image sensor provided 12 bits/pixel images of each spectral band geotagged with the GPS information and pitch, roll, and yaw information from the inertial measurement unit sensor.

Because different UAVs were used each year, the flight conditions varied. For the 3DR Solo used in 2018, the spatial resolution, flight speed, front overlap rate, and side overlap rate were about 1.4 cm at an altitude of 20 m, 2 m/s, 80%, and 80%, respectively. For the M600 used in 2019 and 2020, these were about 3.5 cm at an altitude of 50 m, 3 m/s, 70%, and 70%, respectively. Multiple spectral images per flight were acquired under different flight conditions. The multispectral images were acquired 10–16 days (4 October 2018; 1 October 2019; and 5 October 2020) before harvest.

2.3. Image Processing and Analysis

The geotagged images were mosaicked with geometric and radiometric correction in Pix4d Mapper Pro (Pix4d S.A., Prilly, Switzerland). The mosaicked multispectral images were converted into GNDVI images with the equation in Table 2. The reflectance of the rice canopy area for each spectral image was extracted at each sampling position based on the optimal threshold value for minimizing the effects of water and soil from the GNDVI image.

2.4. Regression Analysis

2.4.1. Partial Least Squares Regression

PLSR is a multivariate linear regression method where the least squares method is applied to linear combinations of independent and dependent variables to derive latent variables (LVs) with high covariance. In other words, PLSR proceeds in the direction that describes both dependent and independent variables. PLSR establishes the optimal LVs that maximize the explained variance in the dependent variable from independent variables and that minimize the predicted residual sum of squares (RSS) and mean square error (MSE) [32].

2.4.2. Ridge Regression

RR is a regularized linear regression method that limits the regression coefficient while minimizing RSS and MSE. The advantage of RR over the least squares method is the tradeoff between the bias and variance [33]. If RR is set to a penalty of zero, it simply produces an unbiased least squares predictor, and the variance for the predictor is large. Increasing the penalty decreases the RR coefficient, which causes a slight bias but sharply decreases the variance and MSE. Above a certain penalty, the decrease in variance with increasing penalty slows down, and the bias significantly increases as the coefficient approaches zero [34]. In other words, the penalty determines the strength of the constraint. As the penalty increases, the flexibility of RR decreases, which decreases the variance but increases the bias.

2.4.3. Artificial Neural Network

ANN is a type of ML that mimics the central nervous system. The perceptron is a simple ANN algorithm that receives multiple signals, such as independent variables, and outputs one signal [35]:

$$f(x) = \begin{cases} 1 & wx + b > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The weight, w , acts like neurons in the brain transferring information through electrical signals. Each input signal is an independent variable that is given a unique w . When the sum of the signals exceeds a predetermined threshold in the activation function, the output is 1; otherwise, the output is 0 or -1 . The bias, b , determines the strength of the perceptron activation. However, a perceptron cannot solve nonlinear problems [36]. This difficulty can be overcome by constructing an ANN, which stacks perceptron layers. The ANN comprises an input layer, a hidden layer, and an output layer. The hidden layer (i.e., black box) can produce the optimal predictor by learning the input and output.

2.5. Development of the Regression Models and Mutual Prediction

PLSR-, RR-, and ANN-based prediction models for the rice yield and protein content were developed with six VIs (green normalization difference vegetation index (GNDVI), normalization difference red-edge index (NDRE), chlorophyll index red edge (CIred-edge), difference NIR/Green green difference vegetation index (GDVI), green-red NDVI (GRNDVI), and medium resolution imaging spectrometer terrestrial chlorophyll index (MTCI)) in Table 2 in Python (Python Software Foundation, USA), as shown in Figure 2. The model performances for single-year data (2018, 2019, and 2020) and multiyear data (2018 + 2019, 2018 + 2020, 2019 + 2020, and all years) were evaluated according to the coefficient of determination (R^2), root-mean-square error (RMSE), and relative error (RE).

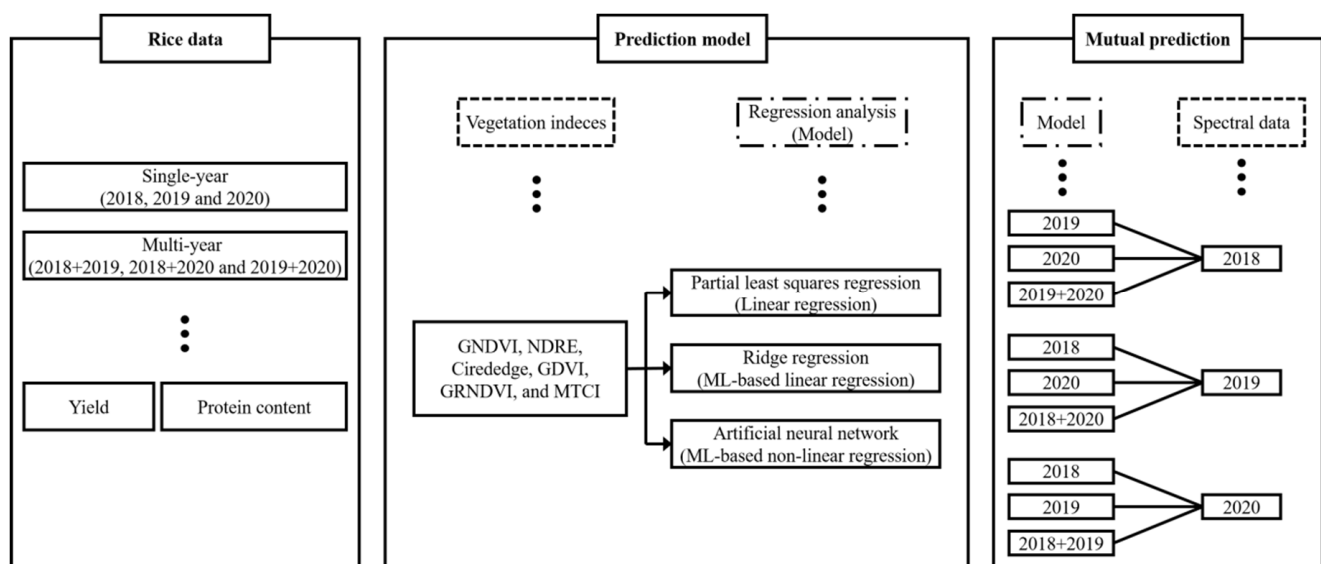


Figure 2. Development of the prediction model and mutual prediction.

Table 2. Vegetation indices calculated from the reflectances of five spectral bands of rice canopies extracted from multispectral images.

Vegetation Index	Equation	Reference
Green normalization difference vegetation index (GNDVI)	$\frac{\rho_{NIR} - \rho_G}{\rho_{NIR} + \rho_G}$	[37]
Normalization difference red-edge index (NDRE)	$\frac{\rho_{NIR} - \rho_{RE}}{\rho_{NIR} + \rho_{RE}}$	[38]
Chlorophyll index red edge (CIrededge)	$\frac{\rho_{NIR} + \rho_{RE}}{\rho_{RE}} - 1$	[39]
Difference NIR/Green green difference vegetation index (GDVI)	$\rho_{NIR} - \rho_G$	[40]
Green-red NDVI (GRNDVI)	$\frac{\rho_{NIR} - (\rho_G + \rho_R)}{\rho_{NIR} + (\rho_G + \rho_R)}$	[41]
Medium resolution imaging spectrometer terrestrial chlorophyll index (MTCI)	$\frac{\rho_{NIR} - \rho_{RE}}{\rho_{RE} + \rho_R}$	[42]

The grid search method was used to identify the best LVs for PLSR, penalty for RR, and epoch for ANN that minimized the MSE. This means that the model exhibited the best performance with the selected tuning parameters. The ANN structure comprised five neurons matching the number of spectral bands as independent variables in a single

hidden layer; this was based on the results of a previous study showing that a single hidden layer is sufficient for predicting the rice-protein content [43]. A rectified linear unit was used to overcome the vanishing gradient problem of sigmoids and serve as the activity function for the ANN [44]. The prediction performances of PLSR, RR, and ANN for the rice yield and protein content were evaluated.

The regression models were cross-validated for single- and multi-year data in nonoverlapping years through mutual prediction, as shown in Figure 2. Cross-validation was used to verify that the model could predict the rice yield and protein content in nonoverlapping years and was evaluated according to the RMSE of prediction (RMSEP) and RE calculated from the 1:1 reference line. Finally, by evaluating the impact of each input VI, the regression model with the highest potential for the year-invariant model was selected.

3. Results

3.1. Rice Reflectance Curves Depending on Nitrogen Fertilization Treatment

Figure 3 shows the average reflectance for each spectral band depending on the amount of N fertilization for each year. Similar to the yield and protein content results, the NIR reflectance reflected the order of the N fertilizer amount. This indirectly means that NIR may be sensitive to N, which affects rice yield and protein content [45,46].

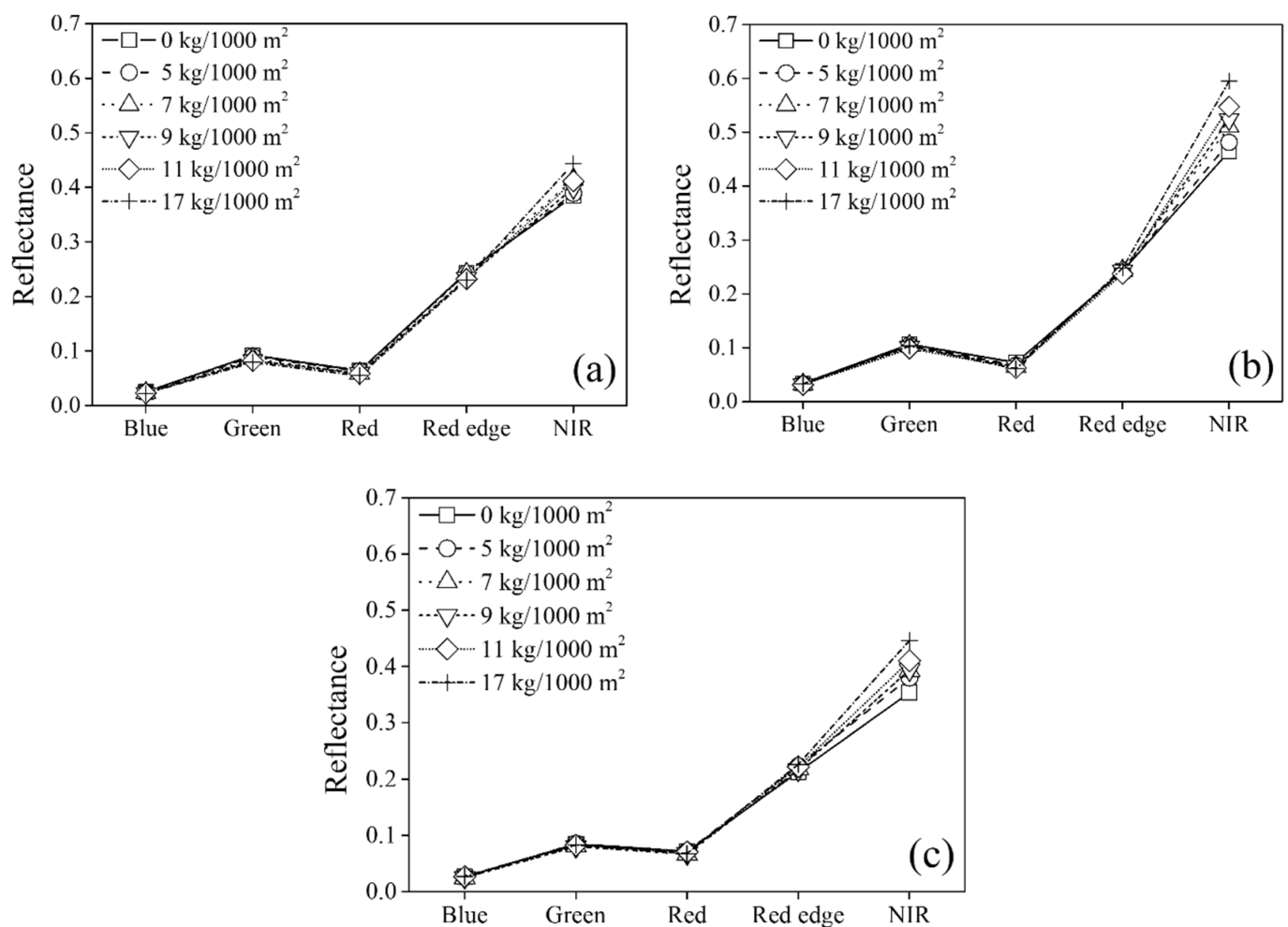


Figure 3. Average reflectance for each spectral band depending on the nitrogen fertilization treatment: (a) 2018, (b) 2019, and (c) 2020.

3.2. Effect of the Nitrogen Fertilization Treatment and Development of Regression Models on the Rice Yield and Protein Content

3.2.1. Rice Yield

Table 3 presents the results of a two-sample t-test with mean and standard deviation (Std) between rice yields depending on the N fertilization treatment in each year. In 2018, the yields did not show significant differences between those obtained via N fertilization treatments of 5, 7, 9, and 11 kg/1000 m². Except for this result, the rice yield increased significantly depending on the amount of fertilizer applied for all years. For all N fertilization treatments, there was no difference between the rice yields in 2018 and 2020. However, the rice yield was higher in 2019 than in the other years.

Table 3. Two-sample t-test among rice yields depending on the nitrogen fertilization treatment for each year.

Nitrogen Treatment	2018 (18 *)	2019 (24)	2020 (24)
0 kg/1000 m ²	419 ± 5.14a **	465 ± 18.7a	415 ± 18.5a
5 kg/1000 m ²	449 ± 9.86b	527 ± 10.3b	433 ± 9.43ab
7 kg/1000 m ²	458 ± 15.0b	546 ± 12.1b	466 ± 22.7b
9 kg/1000 m ²	482 ± 19.0b	591 ± 16.8c	490 ± 24.4c
11 kg/1000 m ²	492 ± 29.9bc	625 ± 7.85d	508 ± 11.9c
17 kg/1000 m ²	534 ± 7.29c	683 ± 16.7e	547 ± 5.58d
All treatment	472 ± 39.9A	573 ± 71.9B	476 ± 47.5A

* Total number of the rice sampling. ** Two-sample t-test at significance level (p -value < 0.05).

Table 4 presents the predicted rice yields with PLSR, RR, and ANN using single-year and multiyear spectral data. The following tuning parameters were selected: 1–4 LVs for PLSR, a penalty of 0.001 for RR, and 844–2372 epochs for ANN. Among the single-year and multiyear analyses in all regression methods, the year combinations of 2019, 2018 + 2019, 2019 + 2020, and all years achieved $R^2 \geq 0.86$. This high performance was attributed to the common inclusion of the 2019 data, which included a wider range of rice yield levels (440–720 kg/1000 m²) than other years. The combinations 2018, 2020, and 2018 + 2020 included the 2018 and 2020 data with a relatively small range of rice yields (380–560 kg/1000 m²). These year combinations yielded $R^2 \geq 0.71$, $RMSE \leq 29.0$ kg/1000 m², and $RE \leq 5.68\%$. Among all single-year and multiyear regression methods, PLSR provided the best prediction performance with $R^2 \geq 0.78$, $RMSE \leq 22.1$ kg/1000 m², and $RE \leq 4.21\%$. RR ($R^2 \geq 0.77$, $RMSE \leq 22.3$ kg/1000 m², and $RE \leq 4.25\%$) provided similar prediction performance with PLSR and better than ANN ($R^2 \geq 0.71$, $RMSE \leq 24.4$ kg/1000 m², and $RE \leq 5.14\%$).

Table 4. Performance of partial least squares regression (PLSR), ridge regression (RR), and artificial neural network (ANN) models in the prediction of rice yield in single- and multi-year analyses.

		R ²	RMSE (kg/1000 m ²)	RE (%)
2018	PLSR (1) *	0.84	15.3	3.24
	RR (0.001)	0.83	15.7	3.32
	ANN (2209)	0.82	20.1	4.25
2019	PLSR (2)	0.91	21.3	3.72
	RR (0.001)	0.91	21.4	3.74
	ANN (2372)	0.89	20.1	3.51
2020	PLSR (1)	0.82	18.9	3.97
	RR (0.001)	0.82	19.0	3.99
	ANN (2341)	0.76	20.2	4.24

Table 4. *Cont.*

		R ²	RMSE (kg/1000 m ²)	RE (%)
2018 + 2019	PLSR (3)	0.93	19.8	3.74
	RR (0.001)	0.93	20.3	3.83
	ANN (2132)	0.88	28.8	5.44
2018 + 2020	PLSR (3)	0.78	18.9	3.98
	RR (0.001)	0.77	18.9	3.98
	ANN (960)	0.71	24.4	5.14
2019 + 2020	PLSR (1)	0.91	22.1	4.21
	RR (0.001)	0.91	22.3	4.25
	ANN (1173)	0.88	26.2	4.99
All years	PLSR (4)	0.91	20.9	4.09
	RR (0.001)	0.91	21.1	4.13
	ANN (844)	0.86	29.0	5.68

* Latent variable in PLSR, penalty in RR, and epochs in ANN used to develop the model with optimum performance.

3.2.2. Rice-Protein Content

Table 5 presents the results of a two-sample t-test with mean and Std between protein contents depending on the N fertilization treatment in each year. For all years, the protein content was either similar under adjacent N fertilization treatment conditions or increased with an increase in the amount of N fertilizer. For all N fertilization treatments, the protein content varied by year. The protein content was highest in 2019, followed by 2020 and 2018. These results show that both the yield and protein content increased with N fertilization. This indicates that applying N fertilization is important for producing high-quality rice with an adequate yield [47].

Table 5. Two-sample t-test among protein contents depending on the nitrogen fertilization treatment for each year.

Nitrogen Treatment	2018 (30 *)	2019 (24)	2020 (24)
0 kg/1000 m ²	5.34 ± 0.10a **	6.20 ± 0.16ab	5.85 ± 0.09a
5 kg/1000 m ²	5.32 ± 0.07a	6.18 ± 0.08a	5.83 ± 0.15ab
7 kg/1000 m ²	5.44 ± 0.14ab	6.35 ± 0.05b	5.88 ± 0.08ab
9 kg/1000 m ²	5.52 ± 0.04bc	6.43 ± 0.15bc	6.03 ± 0.08bc
11 kg/1000 m ²	5.66 ± 0.12c	6.65 ± 0.21cd	6.08 ± 0.08c
17 kg/1000 m ²	6.02 ± 0.07d	6.93 ± 0.15d	6.25 ± 0.05d
All treatment	5.55 ± 0.26A	6.45 ± 0.30B	5.98 ± 0.18C

* Total number of the rice sampling. ** Two-sample t-test at significance level (*p*-value < 0.05).

Table 6 presents the rice-protein contents predicted by PLSR, RR, and ANN based on single-year and multiyear spectral data. The tuning parameters were set as follows: 1–3 LVs for PLSR, a penalty of 0.001 for RR, and 549–2141 epochs for ANN. As stated in Table 5, the protein content was lowest in 2018, intermediate in 2020, and highest in 2019. The linearity was higher in the multiyear regression methods than in the single-year methods because the multiyear analyses merged the different levels of protein content data. Among the single-year and multiyear regression methods, PLSR achieved the highest $R^2 \geq 0.75$ for both single-year and multiyear data, and the errors were $RMSE \leq 0.13\%$ and $RE \leq 2.18\%$. RR ($R^2 \geq 0.74$, $RMSE \leq 0.13\%$, and $RE \leq 2.18\%$) had similar prediction performance to PLSR. ANN provided the lowest prediction performance with $R^2 \geq 0.71$, $RMSE \leq 0.16\%$, and $RE \leq 2.88\%$, except for 2020 ($R^2 = 0.57$, $RMSE = 0.14\%$, and $RE = 2.34\%$). The 2020 model showed the lowest linearity, but similar errors in other years. Therefore, it is necessary for the evaluation of the 2020 model to be made through mutual prediction results. Commonly, the prediction model performance for rice yield and protein content was in the order of PLSR, RR, and ANN.

Table 6. Performance of PLSR, RR, and ANN models in the prediction of rice-protein content in single- and multiple-year analyses.

		R ²	RMSE (%)	RE (%)
2018	PLSR (3) *	0.78	0.11	1.98
	RR (0.001)	0.77	0.11	1.98
	ANN (1514)	0.71	0.16	2.88
2019	PLSR (3)	0.87	0.10	1.55
	RR (0.001)	0.84	0.11	1.70
	ANN (2141)	0.80	0.15	2.32
2020	PLSR (1)	0.75	0.08	1.34
	RR (0.001)	0.74	0.08	1.34
	ANN (2091)	0.57	0.14	2.34
2018 + 2019	PLSR (3)	0.94	0.13	2.18
	RR (0.001)	0.94	0.13	2.18
	ANN (711)	0.91	0.16	2.69
2018 + 2020	PLSR (2)	0.86	0.11	1.92
	RR (0.001)	0.86	0.11	19.2
	ANN (1140)	0.80	0.13	2.26
2019 + 2020	PLSR (1)	0.90	0.10	1.61
	RR (0.001)	0.89	0.11	1.77
	ANN (882)	0.85	0.14	2.25
All years	PLSR (3)	0.93	0.12	2.01
	RR (0.001)	0.93	0.12	2.01
	ANN (549)	0.87	0.16	2.68

* Latent variable in PLSR, penalty in RR, and epochs in ANN used to develop the model with the optimum performance.

3.3. Mutual Prediction

3.3.1. Rice Yield

Table 7 presents the results for the rice yield using PLSR, RR, and ANN models. The prediction performance of each model with single-year and multiyear data was cross-validated against VIs from nonoverlapping years in terms of the RMSEP and RE. Although the PLSR and RR models showed higher prediction performance than the ANN model (see Table 4), the ANN models obtained the lowest mean RMSEPs in the nonoverlapping years 2018 and 2020 (mean RMSEP = 38.5 kg/1000 m² in 2018, mean RMSEP = 29.3 kg/1000 m² in 2020). The PLSR models gave the lowest mean RMSEP in 2019 (mean RMSEP = 41.6 kg/1000 m²). In all single-year and multiyear analyses, the Stds of the RMSEPs were lower in the ANN model (Std RMSEP = 5.61, 10.6, and 4.25 kg/1000 m² in 2018, 2019, and 2020, respectively) than in the PLSR and RR models (Std RMSEP ≥ 14.2, ≥ 14.5, and ≥ 8.62 kg/1000 m² in 2018, 2019, and 2020, respectively). The ANN model, which is an ML-based nonlinear regression, delivered more stable RMSEPs (24.2 kg/1000 m² ≤ RMSEP ≤ 59.1 kg/1000 m²) than the other models (23.7 kg/1000 m² ≤ RMSEP ≤ 83.3 kg/1000 m²) in all single-years and multiyears. The RMSEPs may differ in single-year and multiyear depending on the importance ranking of each VI used in the regression model [48].

Figure 4 plots the mutual rice-yield prediction results of the PLSR and ANN models based on single-year and multiyear data. Note that the predicted and actual results match along the 1:1 line. As shown in Figure 4a, the 2018 PLSR model over-predicted the 2019 and 2020 rice yields (RMSEP ≤ 71.5 kg/1000 m²). The 2018 ANN model also over-predicted the rice yields (RMSEP ≤ 59.1 kg/1000 m²), but to a lesser extent than the PLSR method (Figure 4b). In contrast, the 2019 PLSR and ANN models under-predicted the 2018 and 2020 rice yields (RMSEP ≤ 45.1 kg/1000 m²) (Figure 4c,d). The 2020 ANN model better predicted the 2018 and 2019 rice yields (Figure 4f) than the PLSR model (Figure 4e) (RMSEP = 33.4 kg/1000 m² in ANN 2020 versus ≤ 83.3 kg/1000 m² in PLSR).

In each multiyear, the PLSR model under-predicted the rice yields in nonoverlapping years ($RMSEP \leq 30.2 \text{ kg}/1000 \text{ m}^2$) (Figure 4g). Among the ANN models, the 2019 + 2020 ANN model under-predicted the 2018 rice yields ($RMSEP = 39.1 \text{ kg}/1000 \text{ m}^2$), the 2018 + 2020 ANN model over-predicted the 2019 rice yields ($RMSEP = 42.6 \text{ kg}/1000 \text{ m}^2$), and the 2018 + 2019 ANN model accurately predicted the 2020 rice yields ($RMSEP = 24.2 \text{ kg}/1000 \text{ m}^2$) (Figure 4h). Both the PLSR and ANN models were reproducible with prediction errors of $RMSEP \leq 83.3 \text{ kg}/1000 \text{ m}^2$ and $RE \leq 17.6\%$ in the ripening stage.

Table 7. Mutual prediction of the rice yield via cross-validation with nonoverlapped years using PLSR, RR, and ANN models in single- and multi-year analyses.

		Nonoverlapped Year					
		2018		2019		2020	
		RMSEP (kg/1000 m ²)	RE (%)	RMSEP (kg/1000 m ²)	RE (%)	RMSEP (kg/1000 m ²)	RE (%)
PLSR model	2018	-	-	59.3	10.4	71.5	15.0
	2019	27.9	5.92	-	-	38.8	8.10
	2020	83.3	17.6	41.7	7.29	-	-
	2018 + 2019	-	-	-	-	30.2	6.33
	2018 + 2020	-	-	23.7	4.14	-	-
	2019 + 2020	29.2	6.18	-	-	-	-
	Mean ± Std	46.8 ± 25.8	-	41.6 ± 14.5	-	46.8 ± 17.8	-
RR model	2018	-	-	69.7	12.2	52.6	11.0
	2019	52.8	11.2	-	-	61.2	12.8
	2020	76.1	16.1	33.2	5.79	-	-
	2018 + 2019	-	-	-	-	40.2	8.43
	2018 + 2020	-	-	40.8	7.12	-	-
	2019 + 2020	42.0	8.90	-	-	-	-
	Mean ± Std	57.0 ± 14.2	-	47.9 ± 15.7	-	51.3 ± 8.62	-
ANN model	2018	-	-	59.1	10.3	34.6	7.26
	2019	45.1	9.56	-	-	29.1	6.10
	2020	31.4	6.65	33.4	5.83	-	-
	2018 + 2019	-	-	-	-	24.2	5.08
	2018 + 2020	-	-	42.6	7.44	-	-
	2019 + 2020	39.1	8.29	-	-	-	-
	Mean ± Std	38.5 ± 5.61	-	45.0 ± 10.6	-	29.3 ± 4.25	-

3.3.2. Rice-Protein Contents

Table 8 presents the mutual prediction results of rice-protein content using the PLSR, RR, and ANN models. In the nonoverlapping years 2018 and 2020, the RR and ANN models yielded relatively lower mean RMSEPs ($\leq 0.20\%$ in 2018 and $\leq 0.25\%$ in 2020) than the PLSR model. In 2019, the mean RMSEPs of all PLSR, RR, and ANN models were similar (mean $RMSEP \leq 0.21$). The Std of the RMSEPs were lower in the ANN and RR models (Std $RMSEP \leq 0.04\%$, $\leq 0.08\%$, and $\leq 0.06\%$ in 2018, 2019, and 2020, respectively) than in the PLSR model (Std $RMSEP = 0.24\%$, 0.08% , and 0.28% in 2018, 2019, and 2020, respectively). The RR, and ANN models, which are ML-based regression methods, showed more stable RMSEPs ($0.13\% \leq RMSEP \leq 0.32\%$) than the PLSR model ($0.12\% \leq RMSEP \leq 0.75\%$) in all single-year and multiyear analyses.

Figure 5 graphs the mutual prediction results of the RR and ANN models for the rice-protein content based on single-year and multiyear data. As shown in Figure 5a, the 2018 RR model over-predicted the 2019 and 2020 rice yields ($RMSEP \leq 0.32\%$). Despite poor sensitivity for low rice-protein content in 2020, the predictions of the 2018 ANN model were relatively closer to the 1:1 line ($RMSEP \leq 0.26\%$) (Figure 5b). The 2019 and 2020 ANN models better predicted the 2018 and 2019 rice yields (Figure 5d,f) ($RMSEP \leq 0.21\%$) than the 2019 RR model and the 2020 RR model (Figure 5c,e) ($RMSEP \leq 0.27\%$). The multiyear RR

model better predicted the rice-protein contents in nonoverlapping years ($RMSEP \leq 0.15\%$) (Figure 5g) than the multiyear ANN model ($RMSEP \leq 0.28\%$) (Figure 5h). Both RR and ANN models were reproducible, with prediction errors of $RMSEP \leq 0.32\%$ and $RE \leq 4.97\%$ in the ripening stage.

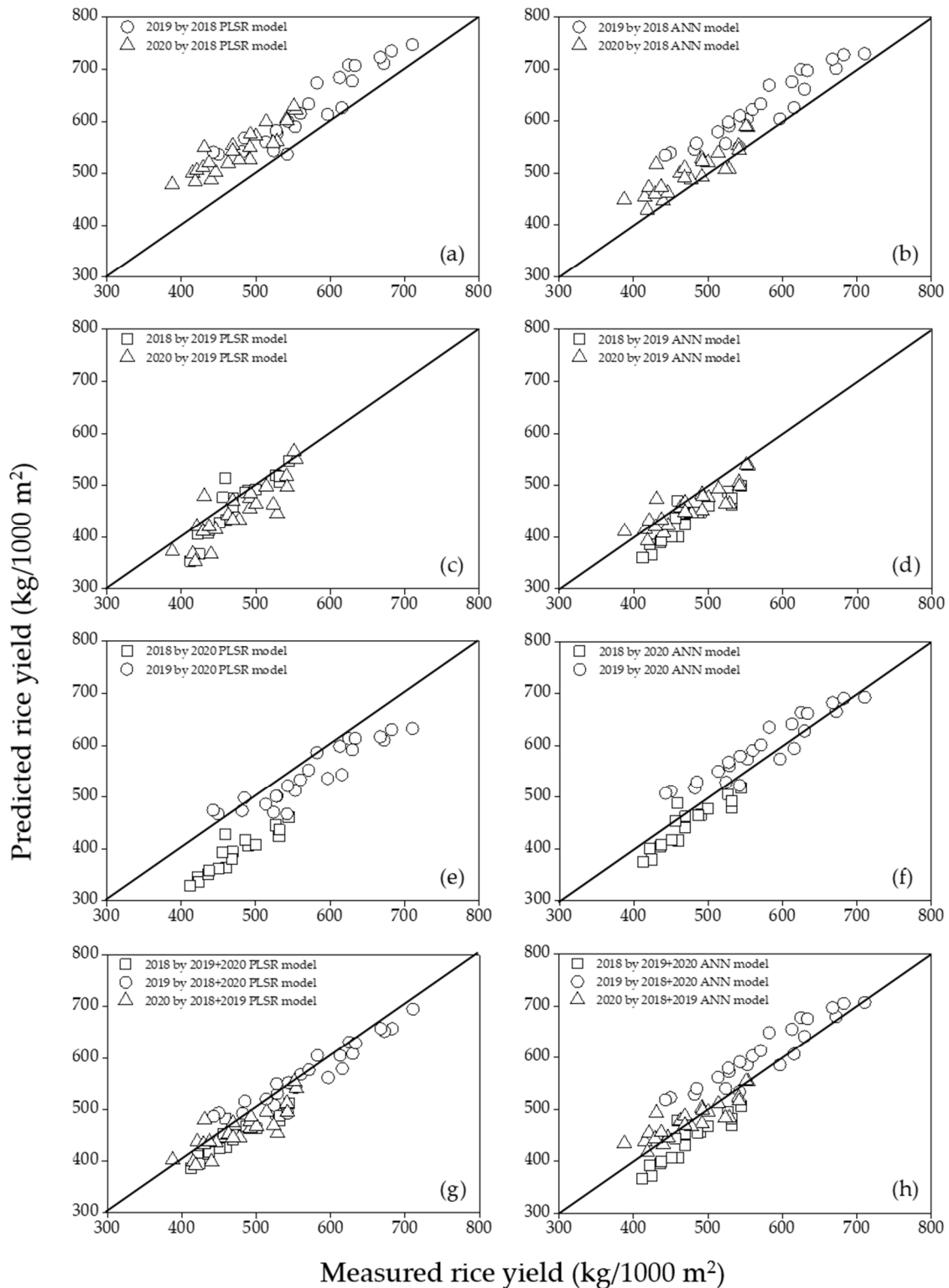


Figure 4. Mutual rice-yield predictions by the PLSR models (a,c,e,g) and the ANN models (b,d,f,h) for single- and multi-year analyses.

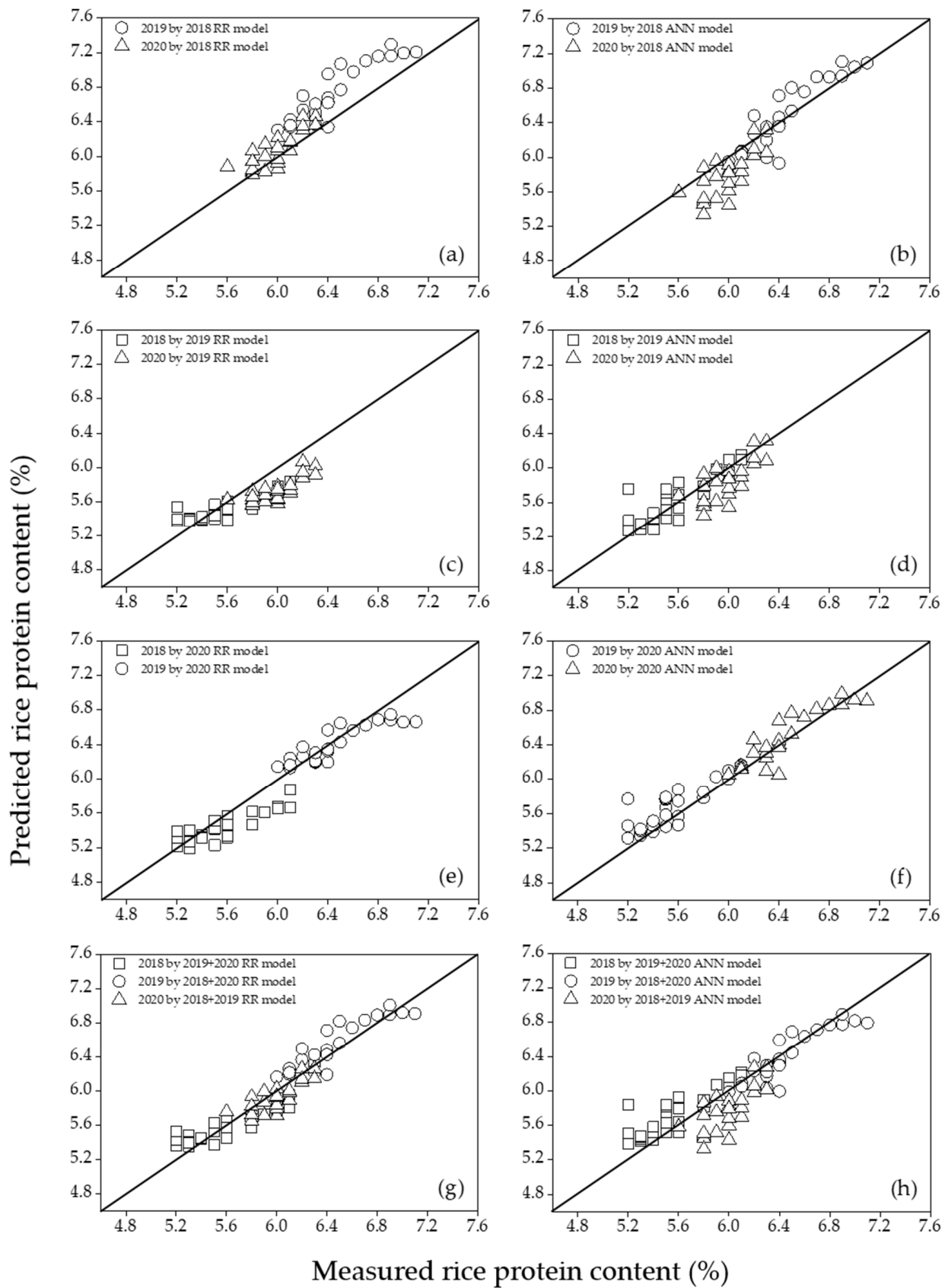
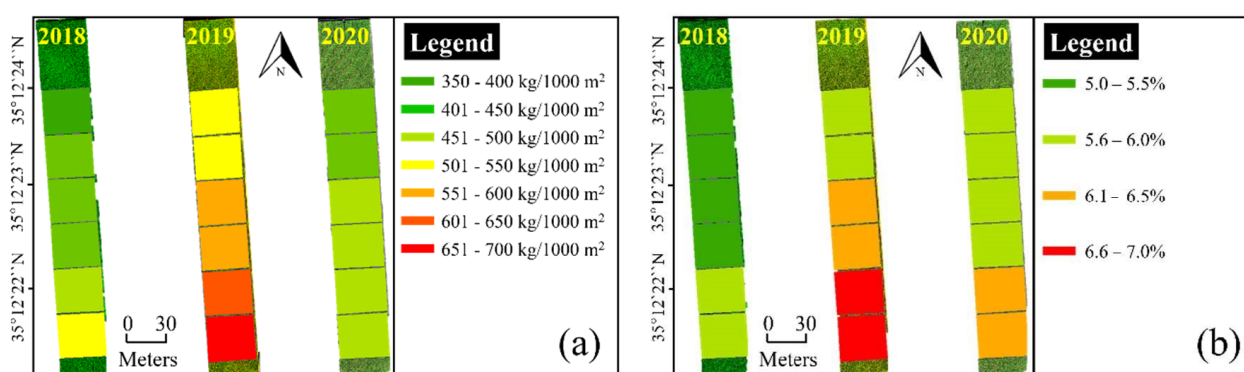


Figure 5. Mutual predictions of the RR models (a,c,e,g) and the ANN models (b,d,f,h) of rice-protein content in single- and multi-year analyses.

Table 8. Mutual prediction of the rice-protein content by cross-validation with nonoverlapped years using PLSR, RR, and ANN models in single- and multi-year analyses.

		Nonoverlapped Year					
		2018		2019		2020	
		RMSEP (%)	RE (%)	RMSEP (%)	RE (%)	RMSEP (%)	RE (%)
PLSR model	2018	-	-	0.33	5.10	0.18	2.99
	2019	0.74	13.3	-	-	0.75	12.5
	2020	0.25	4.57	0.18	2.82	-	-
	2018 + 2019	-	-	-	-	0.12	2.01
	2018 + 2020	-	-	0.13	2.06	-	-
	2019 + 2020	0.20	3.53	-	-	-	-
	Mean ± Std	0.40 ± 0.24	-	0.21 ± 0.08	-	0.35 ± 0.28	-
RR model	2018	-	-	0.32	4.97	0.14	2.42
	2019	0.16	2.97	-	-	0.27	4.57
	2020	0.19	3.39	0.16	2.47	-	-
	2018 + 2019	-	-	-	-	0.13	2.16
	2018 + 2020	-	-	0.15	2.38	-	-
	2019 + 2020	0.13	2.36	-	-	-	-
	Mean ± Std	0.16 ± 0.02	-	0.21 ± 0.08	-	0.18 ± 0.06	-
ANN model	2018	-	-	0.18	2.76	0.26	4.38
	2019	0.16	2.95	-	-	0.21	3.45
	2020	0.19	3.37	0.14	2.23	-	-
	2018 + 2019	-	-	-	-	0.28	4.63
	2018 + 2020	-	-	0.15	2.31	-	-
	2019 + 2020	0.25	4.59	-	-	-	-
	Mean ± Std	0.20 ± 0.04	-	0.16 ± 0.02	-	0.25 ± 0.03	-

Figure 6 maps the rice yield and protein content of each N fertilization treatment field at the ripening stage, predicted by the multiyear ANN models. Figure 6a shows that the rice yield depended on the N fertilization treatment, and the rice yield was predicted to be higher in 2019 than in the other years. The predictions agree with the measured data in Table 3. Figure 6b shows the same pattern for the rice-protein content.

**Figure 6.** Maps of the predicted (a) rice yield and (b) protein content with ANN models for multiple years.

4. Discussion

4.1. Potential for Developing a Year-Invariant Model by Evaluating the Importance of Different Vegetation Indices

In this study, rice plant was grown in the same field for a 3 year period under the same cultivation methods with no environmental disasters. Therefore, it was assumed that the importance ranking of each VI input to the regression model was consistent in each single-year and multiyear. For this reason, each VI should have the same impact on the given data, meaning that the model is potentially available as a year-invariant

model. Applying a regression model with the same rankings of variable importance when predicting rice parameters under different cultivation and environmental conditions will allow relatively easy observation of the changes in prediction performance, depending on the use of input variables.

Figure 7 shows the variance importance in projection (VIP) in the PLSR model, ridge coefficient (RC) in the RR model, and permutation importance (PI) in the ANN model of each VI for predicting rice yield in single years and multiple years. A VIP determines the relative importance of variables; $VIP \geq 1.0$ and $VIP \leq 0.8$ indicate high and low importance, respectively [49]. The greater the difference of the RC from zero, the more important is the variable. The PI method assesses the importance of a variable by its effect on the performance loss when the particular variable is omitted from a black box model [50]. In the rice yield prediction (Figure 7a), the VIPs of all VIs were ≥ 0.88 , but the three most important variables (VIP 1) in each single-year and multiyear were NDRE (2019 + 2020), GDVI (2019, 2018 + 2019, and 2018 + 2020), and MTCI (2018, and 2020). The VIP 2 differed among the single-year and multiyear analyses. Three RC 1s were GDVI (2018 + 2019, and 2019 + 2020), GRNDVI (2019, and 2018 + 2020), and MTCI (2018, and 2020) (Figure 7b). The RC 2 also differed among the single- and multi-year analyses. In contrast, ANN had a single PI 1 variable (MTCI) in all single-year and multiyear analyses (Figure 7c) and a single PI 2 (CIrededge). Together, the MTCI and CIrededge explained more than 80% of all VIs, and may have improved the stability of the RMSEPs ($24.2 \text{ kg}/1000 \text{ m}^2 \leq \text{RMSEP} \leq 59.1 \text{ kg}/1000 \text{ m}^2$) in all single-year and multiyear mutual predictions by ANN (see Table 7). The high importance of these variables might also indirectly explain why the NIR, red edges, and red spectral bands in the MTCI and CIrededge calculations are useful for predicting rice yields. It was considered that the ANN model has the same ranking in the VI of the first and second importance; thus, it best predicts the rice yield in the field.

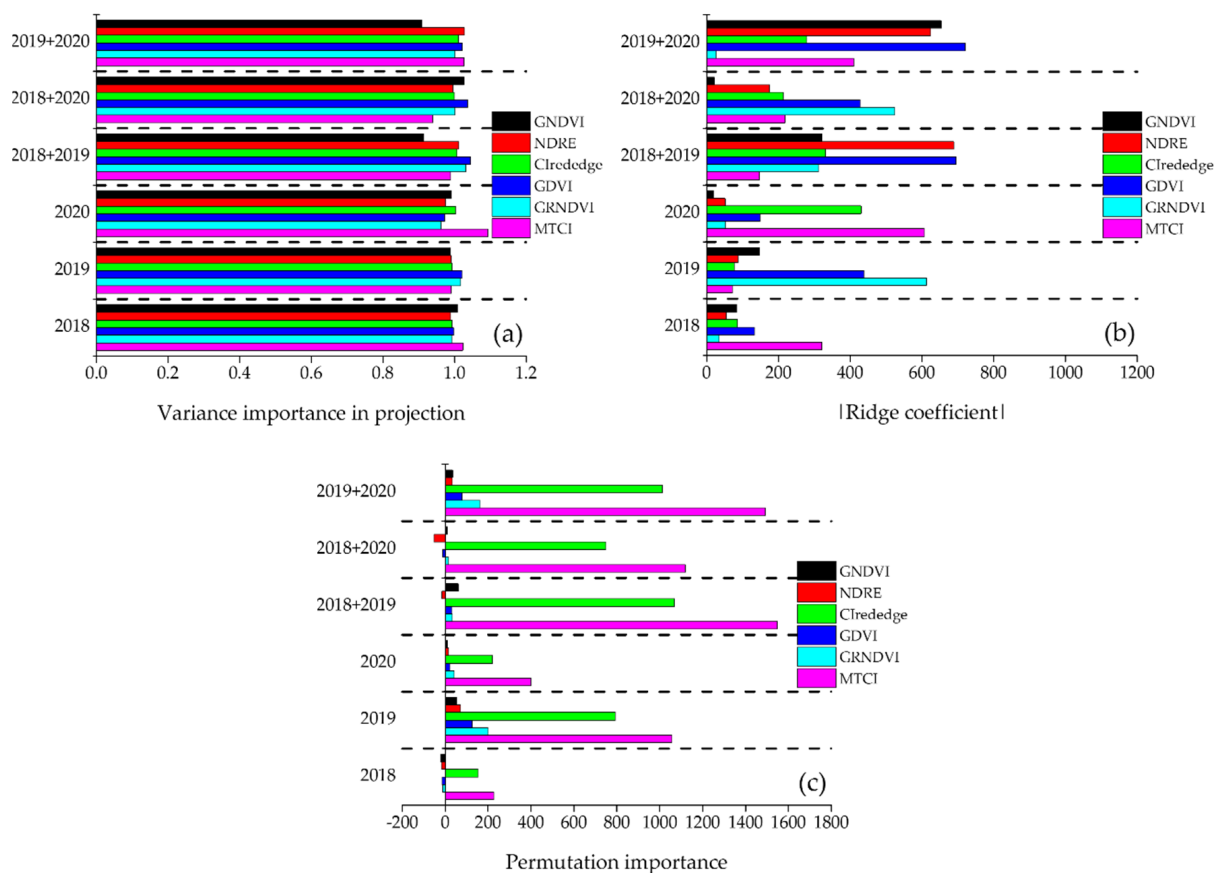


Figure 7. Importance of each vegetation index in (a) PLSR-, (b) RR-, and (c) ANN-based predictions of the rice yield.

Figure 8 shows the VIP in the PLSR model, the RC in the RR model, and the PI in the ANN model of each VI for predicting the rice-protein content in single- and multi-year analyses. Three RC 1s were GNDVI (2019), NDRE (2019 + 2020), or MTCI (2018, 2020, 2018 + 2019, 2018 + 2020) (Figure 8b). The MTCI was ranked VIP 1 and PI 1, except by PLSR in 2019 and by ANN in 2020 (Figure 8a,c). The PLSR and ANN models well explained the rice-protein content, but the ANN model yielded a lower prediction error in mutual prediction than the PLSR model (see Table 8). In 2020, PI 1 and PI 2 were GRNDVI and MTCI, respectively. The GRNDVI was ranked PI 2, except in 2018 + 2019. In conclusion, the ANN model was the most promising year-invariant model. The MTCI and GRNDVI explained more than 71% of all VIs. The MTCI was commonly the most important variable for predicting rice yield and protein content. The second important variables were Clrededge for rice yield and GRNDVI for rice-protein content. For both rice yield and protein content, PI 1 and PI 2 of the multi-year analyses occupied a higher importance ratio than single-year analyses. The importance of a specific variable becomes clearer upon reducing the model complexity of the ANN model; this can help to develop a year-invariant model with increasing reproducibility [20,51].

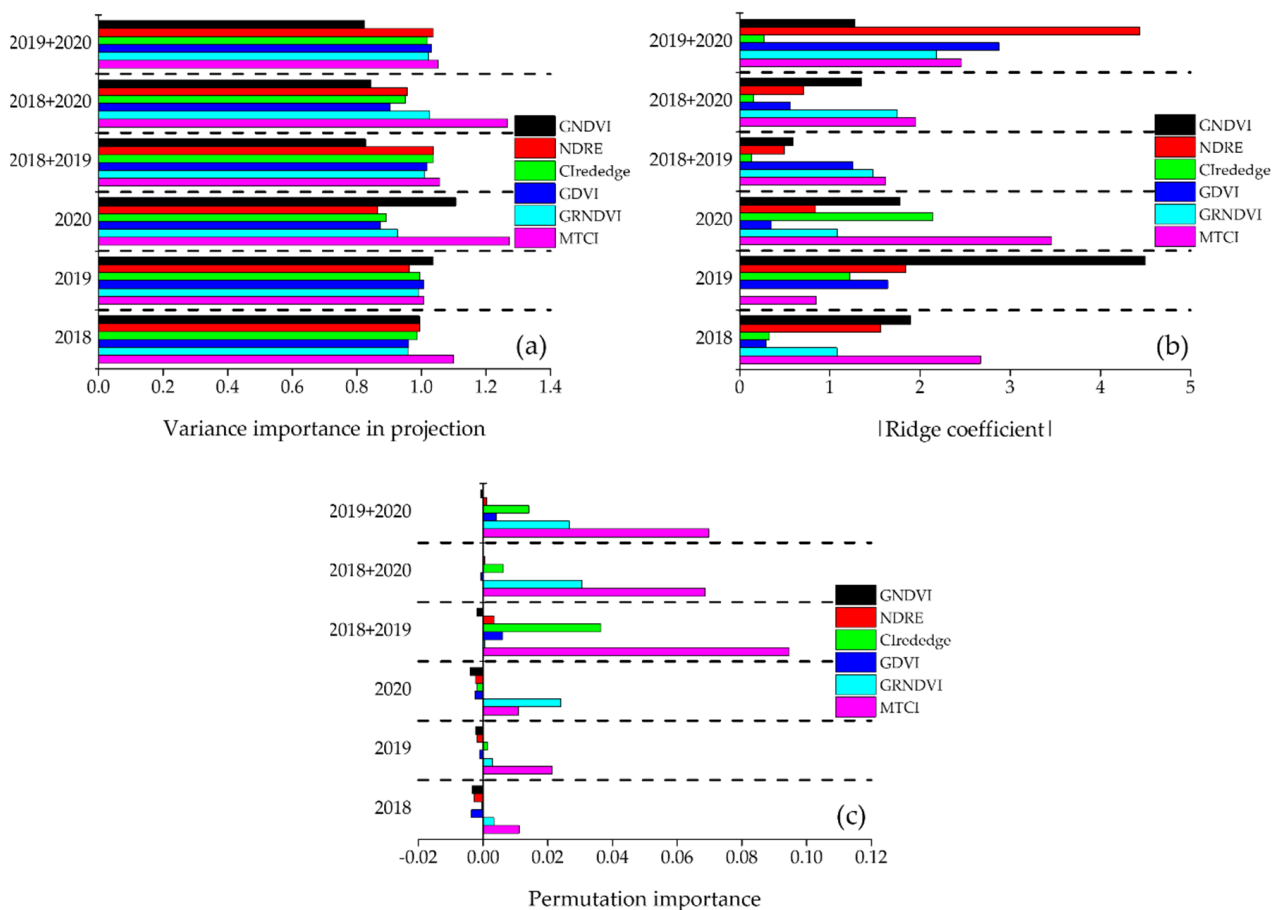


Figure 8. Importance of each vegetation index for the (a) PLSR-, (b) RR-, and (c) ANN-based prediction of the rice-protein content.

4.2. Comparison with and Extension of Related Studies

In previously mentioned studies using linear regression analysis (SLR, MLR, and PLSR), rice yield has been predicted with $R^2 \geq 0.70$ using multiyear data, regardless of the data-acquisition conditions (image sensor, platform, and environmental factor, etc.) [17,21]. However, there are several issues in the prediction of rice-protein content. Most of the studies on the prediction of rice-protein content have been performed using

high-cost hyperspectral image sensors that are disadvantageous for commercialization applications [29,52]. Additionally, the prediction performance varies widely due to other environmental factors, such as shadows caused by clouds and climate conditions. The prediction model shows different slopes and intercepts for each cloud-shadowed area and cloud-free area [23], thus hindering the development of an integrated model for rice-field reproduction. The ANN model, which employs ML-based nonlinear analysis, tried to overcome these issues with multispectral imagery for shadows caused by clouds [43]. As a result, the R^2 of the ANN model was higher (0.92) than that of the PLSR model (0.37). However, there are still limitations; the importance analysis for each input variable has not been performed and it is difficult to perform mutual prediction using single-year data. With reference to the results of previous studies, this study presented the possibility of developing a year-invariant model with analysis of the importance of each vegetation index and mutual prediction results using multiyear data with ANN. Unlike the results in previous studies in which GNDVI was most advantageous in predicting rice-protein content with linear regression [23,24], this study presented MTCI calculated from NIR, red edge, and red as the most important variable. Although the MTCI has been used for prediction of rice-nitrogen content that affects rice-protein content, it has not been reported in rice-protein content prediction with linear regression analysis [53,54]. Therefore, the MTCI may have an important role in predicting rice-protein content using a multispectral image sensor with an ANN model that can overcome environmental factor such as shadows. In addition, different climate conditions for each region affect the predictability of predicting grain yield and protein content. This implies that model calibrations are needed for each cultivation region and year [55]. Some studies have suggested that applying climate data such as temperature, precipitation, and solar radiation as well as important spectral variables such as PI 1 and PI 2 is important to increase predictability on crop yield in other years [28,56]. As a result, when the cumulative temperature data for each year were applied in the mutual prediction of the onion yield model, the RMSEP was lower, with a difference of about 16%. Ultimately, it is necessary to expand this finding to rice fields of various environmental conditions to develop a year-invariant model using the image sensors, modeling methods, and input variables presented in the previous and current study results.

5. Conclusions

PLSR, RR, and ANN were applied to UAV-based multispectral imagery of rice canopies in the ripening stage to develop prediction models for the rice yield and protein content. ANN exhibited a poorer prediction performance ($R^2 \geq 0.71$, $RMSE \leq 29.0$ kg/1000 m², and $RE \leq 5.68\%$ in the case of rice yield and $R^2 \geq 0.57$, $RMSE \leq 0.16\%$, and $RE \leq 2.88\%$ in the case of rice-protein content) than PLSR and RR. However, for an accurate prediction of both rice yield and rice-protein content, the ANN model yielded more stable prediction errors (24.2 kg/1000 m² \leq $RMSEP \leq 59.1$ kg/1000 m² in the case of rice yield, and $0.14\% \leq RMSEP \leq 0.28\%$ in the case of rice-protein content) than PLSR and RR (23.7 kg/1000 m² \leq $RMSEP \leq 83.3$ kg/1000 m² in the case of rice yield and $0.12\% \leq RMSEP \leq 0.75\%$ in the case of rice-protein content) in all single- and multi- analyses. In each analysis, the ANN model gave each VI the same ranking of importance. This consistency may have maintained the RMSEPs in each single-year and multiyear mutual prediction by ANN. For this reason, it was selected as the prediction model that best explained the rice yield and protein content, and as the most promising method for developing a year-invariant model under different cultivation and environmental conditions in the future. The MTCI was commonly the most important variable for predicting rice yield and protein content, followed by Clrededge for rice yield and GRNDVI for rice-protein content. This methodology, which select the optimal model by comparing the ranking of important variables for each year using other regression analysis and by evaluating error stability through mutual prediction, will be useful for the development of available prediction models in the field of agricultural remote sensing. The ANN model proposed in this study can verify the

potential of a year-invariant model by collecting data from the same rice field under similar environmental conditions; however, it cannot verify the reproducibility of the model under different environmental conditions. Therefore, evaluation and reproduction of the ANN model is required for yield and protein content analysis in various environment conditions such as major rice-cultivation complexes.

Author Contributions: Investigation, J.N., S.J., D.S. and Y.K. (Yeseong Kang); resources, S.J., J.N., D.S. and Y.K. (Yeseong Kang); writing—original draft, Y.K. (Yeseong Kang) and C.R.; writing—review and editing, C.R., Y.K. (Younggwang Kim), and S.L.; supervision, C.R. and Y.K. (Younggwang Kim). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Gyeongnam Agricultural Research and Extension Services (Project name: Study on rice growth status monitoring based on unmanned aerial vehicles and Project number: LP003114022020).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors appreciate all staff who helped with this study at the agricultural product research department, Gyeongnam Agricultural Research and Extension Services, Republic of Korea.

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

References

- Inoue, Y.; Moran, M.S.; Horie, T.; Susan, M.M. Analysis of Spectral Measurements in Paddy Field for Predicting Rice Growth and Yield Based on a Simple Crop Simulation Model. *Plant Prod. Sci.* **1998**, *1*, 269–279. [[CrossRef](#)]
- Asseng, S.; Zhu, Y.; Basso, B.; Wilson, T.; Cammarano, D. Simulation Modeling: Applications in Cropping Systems. *Encycl. Agric. Food Syst.* **2014**, *5*, 102–112. [[CrossRef](#)]
- Everingham, Y.; Sexton, J.; Skocaj, D.; Inman-Bamber, G. Accurate prediction of sugarcane yield using a random forest algorithm. *Agron. Sustain. Dev.* **2016**, *36*, 27. [[CrossRef](#)]
- Liu, M.-B.; Li, X.-L.; Liu, Y.; Huang, J.-F.; Tang, Y.-L. Detection of Crude Protein, Crude Starch, and Amylose for Rice by Hyperspectral Reflectance. *Spectrosc. Lett.* **2014**, *47*, 101–106. [[CrossRef](#)]
- Shepherd, K.D.; Walsh, M.G. Infrared spectroscopy—Enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. *J. Near Infrared Spectrosc.* **2007**, *15*, 1–19. [[CrossRef](#)]
- Kang, Y.S.; Ryu, C.S.; Jun, S.R.; Jang, S.H.; Park, J.W.; Song, H.Y.; Sarkar, T.K.; Kim, S.H.; Lee, W.S. Distinguishing between closely related species of Allium and Brassicaceae by narrowband hyperspectral imagery. *Biosyst. Eng.* **2018**, *176*, 103–113. [[CrossRef](#)]
- Fagan, C.C.; Everard, C.D.; McDonnell, K. Prediction of moisture, calorific value, ash and carbon content of two dedicated bioenergy crops using near-infrared spectroscopy. *Bioresour. Technol.* **2011**, *102*, 5200–5206. [[CrossRef](#)] [[PubMed](#)]
- Yang, C.Z.; Shu, X.L.; Zhang, L.L.; Wang, X.Y.; Zhao, H.J.; Ma, A.C.X.; Wu, D.X. Starch Properties of Mutant Rice High in Resistant Starch. *J. Agric. Food Chem.* **2006**, *54*, 523–528. [[CrossRef](#)]
- Chen, C.; McNairn, H. A neural network integrated approach for rice crop monitoring. *Int. J. Remote Sens.* **2006**, *27*, 1367–1393. [[CrossRef](#)]
- Park, C.-E.; Kim, Y.-S.; Park, K.-J.; Kim, B.-K. Changes in physicochemical characteristics of rice during storage at different temperatures. *J. Stored Prod. Res.* **2012**, *48*, 25–29. [[CrossRef](#)]
- Onoyama, H.; Ryu, C.; Suguri, M.; Iida, M. Estimation of rice protein content before harvest using ground-based hyperspectral imaging and region of interest analysis. *Precis. Agric.* **2018**, *19*, 721–734. [[CrossRef](#)]
- Wang, L.; Liu, D.; Pu, H.; Sun, D.-W.; Gao, W.; Xiong, Z. Use of Hyperspectral Imaging to Discriminate the Variety and Quality of Rice. *Food Anal. Methods* **2015**, *8*, 515–523. [[CrossRef](#)]
- Garini, Y.; Young, I.T.; McNamara, G. Spectral imaging: Principles and applications. *Cytom. Part A* **2006**, *69*, 735–747. [[CrossRef](#)]
- Johansen, K.; Raharjo, T.; McCabe, M.F. Using Multi-Spectral UAV Imagery to Extract Tree Crop Structural Properties and Assess Pruning Effects. *Remote Sens.* **2018**, *10*, 854. [[CrossRef](#)]
- Monteiro, S.T.; Minekawa, Y.; Kosugi, Y.; Akazawa, T.; Oda, K. Prediction of sweetness and amino acid content in soybean crops from hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2007**, *62*, 2–12. [[CrossRef](#)]
- Kang, Y.-S.; Kim, S.-H.; Kang, J.-G.; Hong, Y.-K.; Sarkar, T.K.; Ryu, C.-S. Estimation of Leaf Dry Mass and Nitrogen Content for Soybean using Multi-spectral Camera Mounted on Unmanned Aerial Vehicle. *J. Agric. Life Sci.* **2016**, *50*, 183–190. [[CrossRef](#)]
- Huang, J.; Wang, X.; Li, X.; Tian, H.; Pan, Z. Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's-AVHRR. *PLoS ONE* **2013**, *8*, e70816. [[CrossRef](#)]

18. Ryu, C.; Suguri, M.; Umeda, M. Multivariate analysis of nitrogen content for rice at the heading stage using reflectance of airborne hyperspectral remote sensing. *Field Crop. Res.* **2011**, *122*, 214–224. [[CrossRef](#)]
19. Huang, S.; Miao, Y.; Yuan, F.; Gnyp, M.L.; Yao, Y.; Cao, Q.; Wang, H.; Lenz-Wiedemann, V.I.S.; Bareth, G. Potential of RapidEye and WorldView-2 Satellite Data for Improving Rice Nitrogen Status Monitoring at Different Growth Stages. *Remote Sens.* **2017**, *9*, 227. [[CrossRef](#)]
20. Kang, Y.S.; Ryu, C.S.; Kim, S.H.; Jun, S.R.; Jang, S.H.; Park, J.W.; Sarkat, T.K.; Song, H.Y. Yield Prediction of Chinese Cabbage (Brassicaceae) Using Broadband Multispectral Imagery Mounted Unmanned Aerial System in the Air and Narrowband Hyperspectral Imagery on the Ground. *J. Biosyst. Eng.* **2018**, *43*, 138–147.
21. Zhou, X.; Zheng, H.; Xu, X.; He, J.; Ge, X.; Yao, X.; Cheng, T.; Zhu, Y.; Cao, W.; Tian, Y. Predicting grain yield in rice using multi-temporal vegetation indices from UAV-based multispectral and digital imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 246–255. [[CrossRef](#)]
22. Reza, N.; Na, I.S.; Baek, S.W.; Lee, K.-H. Rice yield estimation based on K-means clustering with graph-cut segmentation using low-altitude UAV images. *Biosyst. Eng.* **2019**, *177*, 109–121. [[CrossRef](#)]
23. Ryu, C.; Suguri, M.; Iida, M.; Umeda, M.; Lee, C. Integrating remote sensing and GIS for prediction of rice protein contents. *Precis. Agric.* **2011**, *12*, 378–394. [[CrossRef](#)]
24. Hama, A.; Tanaka, K.; Mochizuki, A.; Tsuruoka, Y.; Kondoh, A. Estimating the Protein Concentration in Rice Grain Using UAV Imagery Together with Agroclimatic Data. *Agronomy* **2020**, *10*, 431. [[CrossRef](#)]
25. Wang, L.; Chang, Q.; Yang, J.; Zhang, X.; Li, F. Estimation of paddy rice leaf area index using machine learning methods based on hyperspectral data from multi-year experiments. *PLoS ONE* **2018**, *13*, e0207624. [[CrossRef](#)] [[PubMed](#)]
26. Campos-Taberner, M.; García-Haro, F.J.; Camps-Valls, G.; Grau-Muedra, G.; Nutini, F.; Crema, A.; Boschetti, M. Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring. *Remote Sens. Environ.* **2016**, *187*, 102–118. [[CrossRef](#)]
27. Sarkar, T.K.; Ryu, C.S.; Kang, J.G.; Kang, Y.S.; Jun, S.R.; Jang, S.H.; Park, J.W.; Song, H.Y. Artificial neural network-based model for predicting moisture content in rice using UAV remote sensing data. *Korean J. Remote Sens.* **2018**, *34*, 611–624.
28. Kang, Y.S.; Jang, S.H.; Park, J.W.; Song, H.Y.; Ryu, C.S.; Jun, S.R.; Kim, S.H. Yield prediction and validation of onion (*Allium cepa* L.) using key variables in narrowband hyperspectral imagery and effective accumulated temperature. *Comput. Electron. Agric.* **2020**, *178*, 105667. [[CrossRef](#)]
29. Onoyama, H.; Ryu, C.; Suguri, M.; Iida, M. Estimation of Rice Protein Content Using Ground-Based Hyperspectral Remote Sensing. *Eng. Agric. Environ. Food* **2011**, *4*, 71–76. [[CrossRef](#)]
30. Nevavuori, P.; Narra, N.; Linna, P.; Lipping, T. Crop Yield Prediction Using Multitemporal UAV Data and Spatio-Temporal Deep Learning Models. *Remote Sens.* **2020**, *12*, 4000. [[CrossRef](#)]
31. Yeo, U.S.; Kim, C.S.; Lee, J.H.; Kwak, D.Y.; Cho, J.H.; Park, D.S.; Song, Y.C.; Shin, M.S.; Yi, G.; Jeon, M.G.; et al. 'Yeonghojinmi': High Grain Quality, Multiple Disease Resistance, and Mid-late Rice Cultivar. *Korean J. Breed. Sci.* **2012**, *44*, 180–184.
32. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
33. Lin, S.B.; Lei, Y.; Zhou, D.X. Boosted Kernel Ridge Regression: Optimal Learning Rates and Early Stopping. *J. Mach. Learn. Res.* **2019**, *20*, 1–36.
34. Algamal, Z.Y. Shrinkage parameter selection via modified cross-validation approach for ridge regression model. *Commun. Stat.-Simul. Comput.* **2020**, *49*, 1922–1930. [[CrossRef](#)]
35. Noriega, L. *Multilayer Perceptron Tutorial*; School of Computing, Staffordshire University: Stoke-on-Trent, UK, 2005.
36. Sagheer, A.; Zidan, M. Autonomous quantum perceptron neural network. *arXiv* **2013**, arXiv:1312.4149.
37. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
38. Maccioni, A.; Agati, G.; Mazzinghi, P. New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra. *J. Photochem. Photobiol. B Biol.* **2001**, *61*, 52–61. [[CrossRef](#)]
39. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [[CrossRef](#)]
40. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
41. Wang, F.-M.; Huang, J.-F.; Tang, Y.-L.; Wang, X.-Z. New Vegetation Index and Its Application in Estimating Leaf Area Index of Rice. *Rice Sci.* **2007**, *14*, 195–203. [[CrossRef](#)]
42. Dash, J.; Curran, P.J. The MERIS terrestrial chlorophyll index. *Int. J. Remote Sens.* **2004**, *25*, 5403–5413. [[CrossRef](#)]
43. Sarkar, T.K.; Ryu, C.S.; Kang, Y.S.; Kim, S.H.; Jeon, S.R.; Jang, S.H.; Park, J.W.; Kim, S.K.; Kim, H.J. Integrating UAV remote sensing with GIS for predicting rice grain protein. *J. Biosyst. Eng.* **2018**, *43*, 148–159.
44. Hinton, G.E. A practical guide to training restricted Boltzmann machines. In *Neural Networks: Tricks of the Trade*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 599–619.
45. Gu, J.; Chen, J.; Chen, L.; Wang, Z.; Zhang, H.; Yang, J. Grain quality changes and responses to nitrogen fertilizer of japonica rice cultivars released in the Yangtze River Basin from the 1950s to 2000s. *Crop J.* **2015**, *3*, 285–297. [[CrossRef](#)]

46. Feng, D.; Xu, W.; He, Z.; Zhao, W.; Yang, M. Advances in plant nutrition diagnosis based on remote sensing and computer application. *Neural Comput. Appl.* **2019**, *32*, 16833–16842. [[CrossRef](#)]
47. Kaur, A.; Ghumman, A.; Singh, N.; Kaur, S.; Viridi, A.S.; Riar, G.S.; Mahajan, G. Effect of different doses of nitrogen on protein profiling, pasting and quality attributes of rice from different cultivars. *J. Food Sci. Technol.* **2016**, *53*, 2452–2462. [[CrossRef](#)] [[PubMed](#)]
48. Anysz, H.; Brzozowski, Ł.; Kretowicz, W.; Narloch, P. Feature Importance of Stabilised Rammed Earth Components Affecting the Compressive Strength Calculated with Explainable Artificial Intelligence Tools. *Materials* **2020**, *13*, 2317. [[CrossRef](#)]
49. Adamowski, J.; Chan, H.F.; Prasher, S.O.; Ozga-Zielinski, B.; Sliusarieva, A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour. Res.* **2012**, *48*, 48. [[CrossRef](#)]
50. Casalicchio, G.; Molnar, C.; Bischl, B. Visualizing the Feature Importance for Black Box Models. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Dublin, Ireland, 10–14 September 2018; Springer: Cham, Switzerland, 2018; pp. 655–670.
51. Das, B.; Nair, B.; Reddy, V.K.; Venkatesh, P. Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *Int. J. Biometeorol.* **2018**, *62*, 1809–1822. [[CrossRef](#)]
52. Xie, X.; Zhang, Y.; Li, R.; Shen, S.; Bao, Y. Prediction model of rice crude protein content, amylose content and actual yield under high temperature stress based on hyper-spectral remote sensing. *Qual. Assur. Saf. Crop. Foods* **2019**, *11*, 517–527. [[CrossRef](#)]
53. Tian, Y.; Yao, X.; Yang, J.; Cao, W.; Hannaway, D.; Zhu, Y. Assessing newly developed and published vegetation indices for estimating rice leaf nitrogen concentration with ground- and space-based hyperspectral reflectance. *Field Crop Res.* **2011**, *120*, 299–310. [[CrossRef](#)]
54. Moharana, S.; Dutta, S. Spatial variability of chlorophyll and nitrogen content of rice from hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *122*, 17–29. [[CrossRef](#)]
55. Wang, L.; Tian, Y.; Yao, X.; Zhu, Y.; Cao, W. Predicting grain yield and protein content in wheat by fusing multi-sensor and multi-temporal remote-sensing images. *Field Crop Res.* **2014**, *164*, 178–188. [[CrossRef](#)]
56. Ma, J.-W.; Nguyen, C.-H.; Lee, K.; Heo, J. Regional-scale rice-yield estimation using stacked auto-encoder with climatic and MODIS data: A case study of South Korea. *Int. J. Remote Sens.* **2019**, *40*, 51–71. [[CrossRef](#)]