

## Article

# Analysis of the Effects of Population Structure and Environmental Factors on Rice Nitrogen Nutrition Index and Yield Based on Machine Learning

Yan Jia <sup>1,†</sup>, Yu Zhao <sup>2,\*,†</sup> , Huimiao Ma <sup>1</sup>, Weibin Gong <sup>1</sup>, Detang Zou <sup>1</sup>, Jin Wang <sup>3,4</sup>, Aixin Liu <sup>1</sup>, Can Zhang <sup>1</sup>, Weiqiang Wang <sup>1</sup>, Ping Xu <sup>1</sup>, Qianru Yuan <sup>1</sup>, Jing Wang <sup>1</sup>, Ziming Wang <sup>1</sup> and Hongwei Zhao <sup>1,\*</sup> 

<sup>1</sup> Key Laboratory of Germplasm Enhancement, Physiology and Ecology of Food Crops in Cold Region, Ministry of Education, Northeast Agriculture University, Harbin 150030, China; jiajan\_cool@126.com (Y.J.); ma2213097688@163.com (H.M.); g314159261201@163.com (W.G.); zoudt@163.com (D.Z.); 18724510883@163.com (A.L.); zc445496019@126.com (C.Z.); wwq20012024@163.com (W.W.); xp2831519695@163.com (P.X.); 13467078458@163.com (Q.Y.); 19862024614@163.com (J.W.); 13679336508@163.com (Z.W.)

<sup>2</sup> College of Electronic and Information, Northeast Agricultural University, Harbin 150030, China

<sup>3</sup> State Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Nanjing Agricultural University, Nanjing 210095, China; wangjin\_cool@163.com

<sup>4</sup> Bei Da Huang Kenfeng Seed Limited Company, Harbin 150431, China

\* Correspondence: yuzhao@neau.edu.cn (Y.Z.); hongweizhao\_cool@126.com (H.Z.)

† These authors contributed equally to this work.

**Abstract:** With the development of rice varieties and mechanized planting technology, reliable and efficient nitrogen and planting density status diagnosis and recommendation methods have become critical to the success of precise nitrogen and planting density management in crops. In this study, we combined population structure, plant shape characteristics, environmental weather conditions, and management information data using a machine learning model to simulate the responses of the yield and nitrogen nutrition index and developed an ensemble learning model-based nitrogen and planting density recommendation strategy for different varieties of rice types. In the third stage, the NNI and yield prediction effect of the ensemble learning model was more significantly improved than that of the other two stages. The scenario analysis results show that the optimal yields and nitrogen nutrition indices were obtained with a density and nitrogen amount of  $100.1 \times 10^4$  plant/ha and  $161.05 \text{ kg} \cdot \text{ha}^{-1}$  for the large-spike type variety of rice,  $75.08 \times 10^4$  plant/ha and  $159.52 \text{ kg} \cdot \text{ha}^{-1}$  for the intermediate type variety of rice, and  $75.08 \times 10^4$  plant/ha and  $133.47 \text{ kg} \cdot \text{ha}^{-1}$  for the panicle number type variety of rice, respectively. These results provide a scientific basis for the nitrogen application and planting density for a high yield and nitrogen nutrition index of rice in northeast China.

**Keywords:** nitrogen and density management; ensemble learning model; nitrogen nutrition index



**Citation:** Jia, Y.; Zhao, Y.; Ma, H.; Gong, W.; Zou, D.; Wang, J.; Liu, A.; Zhang, C.; Wang, W.; Xu, P.; et al. Analysis of the Effects of Population Structure and Environmental Factors on Rice Nitrogen Nutrition Index and Yield Based on Machine Learning. *Agronomy* **2024**, *14*, 1028. <https://doi.org/10.3390/agronomy14051028>

Academic Editor: Gniewko Niedbała

Received: 7 April 2024

Revised: 30 April 2024

Accepted: 9 May 2024

Published: 12 May 2024



**Copyright:** © 2024 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

One of the biggest challenges facing global agriculture is feeding the growing population, as the world population will require an approximately 70% increase in food production by 2050 [1]. Rice (*Oryza sativa* L.), a staple food for more than half of the world's population, is vital to ensuring China's food security [2,3]. Global rice demand for food security is forecasted to reach 116 million tons by 2035 [4]. However, with the continuous expansion of urbanization and industrialization in the 21st century, the area of arable land has been declining year-by-year [2]. Therefore, increasing rice yield per unit area has become a reliable option to address the above challenges [5]. Hence, an increase in rice productivity per unit area seems to be a viable option to meet the aforementioned challenges [6].

The nitrogen (N) application amount and transplanting density are important field management factors that affect the rice yield, and the optimum N fertilizer rate and

transplanting density are different for different varieties of rice [2,7]. As a decision support tool for agricultural systems, precision agriculture can maximize crop yield and quality, ensure a steady increase in resource utilization efficiency and agricultural production efficiency, and maximize sustainable agricultural development [5,8]. At present, most of the relevant research on precision agriculture models has focused on the effect of N fertilizer application on crop yield [5,9]. Precision N management (PNM) can be used to systematically monitor the relationship between the N supply and plant N demand in time and space and improve N use efficiency (NUE) by formulating a reasonable amount of N fertilizer [9]. However, there have been few studies related to precision agriculture models on the effects of N application and density on rice yield, and these have mainly focused on the analysis of data collected by remote sensing technology [10]. There has been no systematic analysis of characteristic data, such as population structure characteristics and plant type characteristics.

In China, most farmers choose to increase the amount of N fertilizer to achieve a high crop yield [11]. In particular, excessive N fertilizer application not only reduces N recovery efficiency, but also causes environmental pollution, such as runoff, denitrification, leaching, and volatilization, which increase N loss [7]; it also causes plant lodging and yield loss. Therefore, finding the appropriate N fertilizer dosage for different varieties is one of the main ways to achieve a high yield and high efficiency of rice [12].

N accumulation-related diagnostic indexes [13] and N concentration-related indicators are used to diagnose crop growth and the N status of crops [14,15]. The N nutrient index (NNI) is defined as the ratio of the actual plant N concentration ( $N_a$ ) to the critical plant N concentration ( $N_c$ ). The NNI is a reliable indicator for diagnosing the N status of crops. In recent years, some studies have established leaf area index (LAI)-based N diagnostic models [16]. In addition, the efficient use of fertilizers in the field is closely related to the fertilizer type and soil properties (e.g., soil temperature, soil moisture, soil pH, and soil aeration) [10]. However, there is still a lack of relevant model research to comprehensively predict the NNI and then diagnose yield based on factors such as the population structure, crop variety, fertilizer dosage, planting density, field soil environment, field meteorological resources, and environment.

Plant density directly affects the number of effective panicles by affecting the development of tillers [17], and different types of rice varieties have different suitable planting densities [18,19]. Moreover, because mechanical transplanting has gradually replaced manual transplanting [20], technical measures for N reduction with high transplanting density have become low-carbon and environmentally friendly rice cultivation technical measures, which not only help to increase plant density and reduce N, but also help to save on labor costs, thus producing considerable economic benefits [21]. Therefore, formulating N fertilizers and planting densities suitable for different types of rice may be one of the key ways to achieve high yields and high NUE in modern rice production systems.

Statistical methods were used to evaluate crop models so as to predict the yield capacity of major crops based on meteorological data, yield statistics, and reproduce yield variability, as well as to determine the agricultural processes that affect the yield of major European crops [22]. It is not easy to predict yields at regional scales using crop models, and the yields predicted by models are not more accurate than those predicted by simple regression models [23,24].

Over the past decade, owing to their ability to extract information hidden in data, machine learning (ML) methods have developed rapidly in the field of precision agriculture, providing a wide range of effective solutions for crop and environmental state estimation and decision support [5,10,25–29]. They can solve significant nonlinear problems using datasets from multiple sources. Moreover, the combination of different ML methods and signal processing technology can make great contribution for precision agriculture in the future [30]. ML and deep learning methods were used to predict rice yields in large areas using multi-source remote sensing data [31]. Random forest regression (RFR) and a simple linear regression model of ML were used to predict corn yield and the NNI,

combining soil, weather, and management information, and it was found that the RFR model better predicted the corn NNI and grain yield ( $R^2 = 0.86$  and  $0.79$ ) [5]. The hybrid method, combines ML regression with a physically based approach to evaluate the crop N content. A Gaussian process-based sequential backward band removal algorithm was used to analyze the information content of specific frequency bands in a leaf optical model simulated spectrum to estimate the aboveground N, and the root mean square error (RMSE) was  $2.1 \text{ g/m}^2$  [32]. While environmental, genetic, and management data were used together with active canopy sensor data, support vector regression ( $R^2 = 0.74$ – $0.90$  for prediction) or the RFR model ( $R^2 = 0.84$ – $0.93$  for prediction) could be used to predict N status indicators of maize more reliably, compared with using the normalized difference vegetation index (NDVI), or normalized difference red edge [29]. RFR models combining weather, soil, and plant growth characteristics with spectral indices showed a highly improved predictive reliability ( $R^2 = 0.81$ ) compared with models based on NDVI, and an N rate of  $150 \text{ kg N ha}^{-1}$  was recommended for most canola production [28].

However, there have been few related studies on the prediction of a suitable planting density and N fertilizer dosage for rice based on ensemble learning models [33]. Most of the current models focus on the performance of the crop NNI and yield prediction models and recommendation algorithms for N fertilizer, and these were reported to be influenced by soil properties [34], environmental weather conditions [5], field management techniques [35], and crop variety. Therefore, it is important to effectively combine the above factors to improve the current crop NNI and yield prediction and N recommendations [36]. Most of the data used by ML models in the past were obtained by sensors [12,25,28], and were not combined with characteristic value data such as the population structure, plant type traits, field meteorological data, and N fertilizer and density for prediction of the yield and NNI. Therefore, there is an urgent need to develop a PNM strategy combined with plant population structure and plant shape characteristics for sustainable rice production in northeast China. Currently, the regional optimum N management strategy promoted by extension services recommends a total N content of about  $180 \text{ kg N ha}^{-1}$  for rice [37].

At present, most of the PNM model predictions about the NNI and yield level use the RFR model [5,28,29,38]. The ensemble learning method of the RFR model has a strong generalization performance and reduces overfitting when dealing with different data set segmentations [27]. However, there have been few related studies on the prediction of the yield and NNI using the Light Gradient Boosting Machine (LGBM) model focusing on the ensemble learning method. As the LGBM model is an efficient gradient boosted decision tree, it speeds up the training process and is sometimes more accurate [26]. To solve this problem, we aim to use machine learning methods to find out the optimal transplanting density and nitrogen fertilizer for different types of varieties of rice.

Therefore, our study employed two ensemble learning methods (RFR and LGBM) to predict the yield and NNI. The objectives of this study were as follows: (1) improve prediction of the NNI and yield by combining the population structure, plant shape characteristics, environment, and nutrient information data using ensemble learning models (RFR and LGBM) as compared with the ridge regression model (RR) using similar data; (2) according to the results of the ensemble learning models, we analyzed the main controlling factors affecting the yield and NNI, and the relationship between yield and NNI and the main controlling factors among different types of varieties; (3) develop an N fertilizer and density recommendation for different rice varieties based on the ensemble learning models using the population structure, plant shape characteristics, environment, and nutrient information data.

## 2. Materials and Methods

### 2.1. Plant Material and Growth Conditions

This study was conducted from the year 2014 to 2021 in XiangFang County (longitude:  $126^{\circ}74'$ – $126^{\circ}76'$  E; latitude:  $45^{\circ}71'$ – $45^{\circ}73'$  N), Acheng County (longitude:  $126^{\circ}22'$ – $126^{\circ}50'$  E; latitude:  $45^{\circ}34'$ – $45^{\circ}46'$  N), and the Northeast Agriculture University (longitude:  $126^{\circ}71'$  E;

latitude: 45°7' N) in the Heilongjiang Province in northeast China (Figure S1). Three study sites were selected, and the soil properties are shown in Table 1. We collected 10–20 cm of soil in the tillage layer for measurement before transplanting every year. Indicators in Table 1 were determined using the soil base fertility methods [39]. Site 1, Site 2, and Site 3 were black soil (loamy clay) equivalent to typic Haploboroll, according to the United States Department of Agriculture (USDA) Soil Taxonomy. The design of the experiments conducted at the three study sites are shown in Table 2.

**Table 1.** The soil properties at the three study sites during 2014–2021.

Site	Soil Type	Soil Texture	Year	Organic Matter (g/kg)	Total Nitrogen (g/kg)	Total Phosphorus (g/kg)	Slowly Available K (mg/kg)	Available N (mg/kg)	Available P (mg/kg)	Available K (mg/kg)	Value of pH
Site 1	Black soil	Loamy clay	2014	22.30	1.20	0.40	706.5	129.8	18.70	99.1	6.80
			2015	22.21	1.24	0.41	706.2	128.9	18.52	99.3	6.70
			2016	22.34	1.21	0.39	705.2	126.4	18.60	98.6	6.79
			2017	22.13	1.12	0.38	704.2	126.4	25.31	97.9	6.42
Site 2	Black soil	Loamy clay	2018	22.18	1.19	0.34	695.9	127.8	18.80	85.7	6.82
			2019	22.05	1.18	0.44	704.2	125.4	17.97	92.4	6.62
			2019	34.89	1.51	0.94	708.7	130.5	20.6	91.4	6.56
Site 3	Black soil	Loamy clay	2020	32.16	1.62	0.86	706.2	128.6	19.8	90.4	6.23
			2021	33.13	1.54	0.90	702.4	131.3	18.6	89.4	6.38

**Table 2.** The design of the experiments conducted at the three study sites.

Site	Year	Planting Density (10,000 Plants/ha)	Nitrogen Fertilizer Application (kg/ha)	Variety
Site 1	2014, 2015, 2016	180.18, 150.15, 128.70, 120.12, 100.10, 90.09, 85.80, 75.08, 72.07, 64.35, 60.06, 51.48	0, 75, 150, 225	DN425, DN427, DN426, NDJ30, SJ14
Site 2	2017, 2018, 2019	60.06, 75.08, 100.10	0, 75, 150, 225	DN426, MDJ30, SJ14
Site 3	2019, 2020, 2021	75.08	150	SJ9, SJ18, SJ21, SJ3, T256

Note: Site 1 was located in XiangFang County (longitude: 126°74'–126°76' E; latitude: 45°71'–45°73' N), Site 2 was located in Acheng County (longitude: 126°22'–126°50' E; latitude: 45°34'–45°46' N), and Site 3 was located at the Northeast Agriculture University (longitude: 126°71' E; latitude: 45°7' N) in the Heilongjiang Province in northeast China.

The experimental materials used were ten japonica rice cultivars, DN425 and SJ14 (a low-tillering, large-sized panicle cultivar, large-spike type variety), DN427 and NDJ30 (a relatively heavy-tillering, small-sized panicle, panicle number variety), and DN426, SJ9, SJ18, SJ21, SJ3 and T256 (a middle-tillering, mid-sized panicle cultivar, intermediate type variety). The characteristics of the rice varieties refer to the National Rice Data Center [40] and laboratory pre-field multi-year measurements of yield and yield components.

## 2.2. Experimental Design

A summary of weather data between 2014 and 2021 and crop duration date are shown in Table 3; we have set up a total of 843 plots from 2014 to 2021. For Site 1, the experiment was conducted from 2014 to 2016, with three replications. The experiment employed a randomized complete block design with split plots, including twelve planting densities (D1: 180.18, D2: 150.15, D3: 128.70, D4: 120.12, D5: 100.10, D6: 90.09, D7: 85.80, D8: 75.08, D9: 72.07, D10: 64.35, D11: 60.06, D12: 51.48,  $10^4$  plants/ha) as main plots, and four N fertilizer applications (N0: 0, N1: 75, N2:150 and N3: 225, kg N/ha) as sub-plots (with a plot size of 150 m<sup>2</sup>). Four varieties were selected for Site 1. The number of plots was 240 per year from 2014 to 2016, and the cumulative number of plots in the 3 years was 720. For Site 2, the experiment was conducted from 2017 to 2019, with three replications. The experiments employed a randomized complete block design with split plots, including three planting densities (D1: 60.06, D2: 75.08, D3: 100.10,  $\times 10^4$  plants/ha) as main plots, and four N fertilizer rates (N0: 0, N1: 75, N2:150 and N3: 225, kg N/ha) as sub-plots (with

a plot size of 150 m<sup>2</sup>). Three varieties were selected for Site 2. The number of plots was 36 per year from 2017 to 2019, and the cumulative number of plots in the 3 years was 108. Meanwhile, for Site 3, from 2019 to 2021, there were three replications. The planting density was  $75.08 \times 10^4$  plant/ha and the N fertilizer rate was 150 kg/ha, arranged in a randomized complete block design with three replications (with a plot size of 100 m<sup>2</sup>). Five varieties were selected for Site 3. The number of plots was 5 per year from 2019 to 2021, and the cumulative number of plots in the 3 years was 15. For each plot, sufficient phosphate (75 kg P<sub>2</sub>O<sub>5</sub> ha<sup>-1</sup>) and potash (50 kg K<sub>2</sub>O ha<sup>-1</sup>) fertilizers were applied before planting to make sure that the phosphorus and potassium nutrients were not limiting the plant growth. All plots were kept free of weeds, insects, and diseases with pesticide applications based on local standard practices.

**Table 3.** A summary of weather data between 2014 and 2021. Ten different rice varieties during 2014 and 2021 in this table were used in the analysis.

Continuous Variables	Level/Unit	Mean	Standard Deviation
Average maximum temperature during March and October	°C	24.84	3.55
Average minimum temperature during March and October	°C	14.33	4.30
Monthly average of cumulative radiation during March and October	MJ/m <sup>2</sup>	503.37	34.59
Crop duration	Days	138	2.98

### 2.3. Field Sampling and Data Collection

The population structure and plant shape characteristics data were collected at full heading stage, and NNI and grain yield was determined at the full maturity stage. The soil information was obtained before transplanting and the weather data was measured throughout the whole growth period.

At the peak of leaf area (heading stage), 3 hills of representative plants were selected from each plot, and the punching and weighing method was adopted (a circular puncher with a diameter of 3 cm<sup>2</sup> was used to punch holes in the upper, middle, and lower parts of the leaves, and we dried the small discs of the leaves and weighed them, calculated the leaf area, and repeated the method three times). The leaf area was calculated as follows:

$$\text{Leaf area} = W \times A \quad (1)$$

where  $W$  is the dried weight of the small discs of the leaves and  $A$  is the area of the circular puncher.

At the full heading stage, we used a Chlorophyll Meter Model SPAD-502 to measure the living leaves, selected and measured, in turn, the mid-section and about 3 cm up and down the middle-section of the flag leaves on the main stem of the plant, and calculated the average value to represent the SPAD value (avoiding leaf veins when measuring). The main stems of 5 hills of representative plants were selected from each plot, and this was repeated 3 times.

During the full heading stage, on sunny days from 11:00 to 13:00, we used the plant canopy analyzer AccuPARLP-80 from the Decagon Company in the United States. We obtained the determination of photosynthetically active radiation (PAR) in rice. The external sensor was fixed on the top of the 2 m straight rod to measure the PAR at the top of the canopy, and then the horizontal probe was used to measure the PAR at the bottom of the canopy, and the extinction coefficient  $K$  was calculated using Beer's law.

$$\text{Extinction coefficient}(K) = -\frac{\ln Tr}{LAIc} \quad (2)$$

where  $Tr$  is the light transmittance of the canopy group and  $LAIc$  is the leaf area index of the rice group measured by the instrument, not the actual value.

At the full heading stage, 5 representative points with consistent growth were selected for each treatment, and each was investigated once. When measuring, the canopy width was the distance between tillers that were the farthest apart from the ground at a certain height. The plant height was investigated at the full heading stage, and the plant height was the height of the plant from the base to the highest point of the panicle. At the full heading stage, the included angle of the flag leaves was measured, and the included angle of the leaves of 10 plants with the same growth vigor was measured with a protractor, which was repeated 10 times.

The plant aboveground biomass was selected from three representative plants in each plot. The stem, leaf, and spike were separated, and dried at 105 °C for 30 min, then dried at 80 °C for constant weight. A digestion method was selected for the determination of N content in the plant tissues [41]. The NNI was calculated as follows [15]:

$$\text{NNI} = N_a/N_c \quad (3)$$

where  $N_a$  was the actual plant N concentration and  $N_c$  was the critical plant N concentration.

At maturity, 2 square meters plots from the center of each subplot were harvested for the yield determination and the results were normalized to 14% grain moisture content.

#### 2.4. Regression Models for Predicting the Rice NNI and Yield

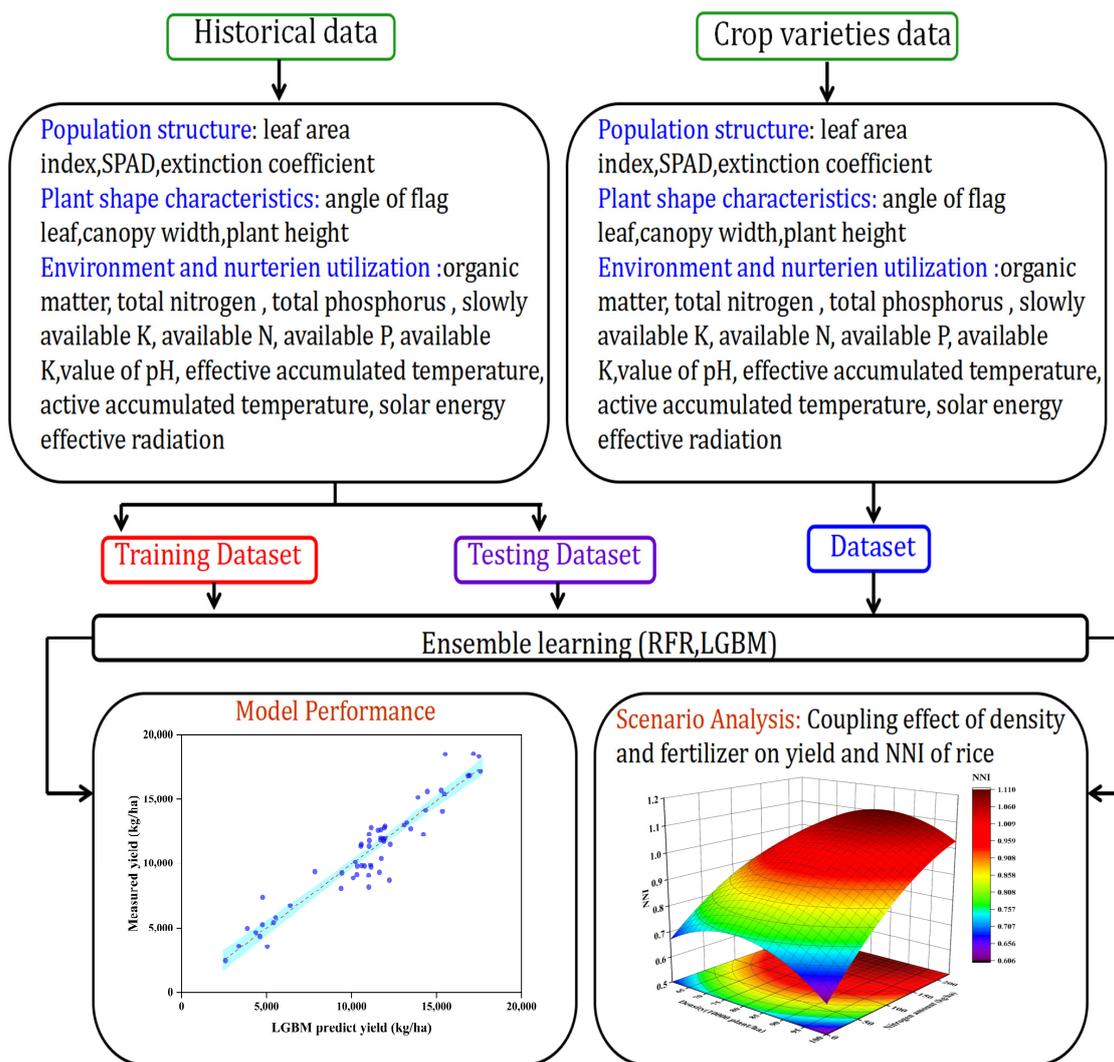
In this study, linear regression and ensemble learning models were used. Ridge regression (RR) was carried out as a linear regression. Random forest regression (RFR) and Light Gradient Boosting Machine (LGMB) models were used as ensemble learning methods in machine learning. All models were adopted to predict NNI and yield in 3 stages.

In the first stage, the models were used to predict the rice yield and NNI with population structure, which was leaf area index, SPAD value, and extinction coefficient. In the second stage, the models were used to predict the rice yield and NNI with population structure and plant shape characteristics, and the shape characteristics included base angle of flag leaf, canopy width, and plant height. In the third stage, environment and nutrient utilization were added to predict the rice yield and NNI, like organic matter, total nitrogen, total phosphorus, slowly available K, available N, available P, available K, value of pH, effective accumulated temperature, active accumulated temperature, and solar energy effective radiation (RA).

For the RR, the parameter of alpha was adjusted, and the best value was 0.2 for the NNI prediction, and 0.1 for the yield prediction, respectively. For the RFR model, the parameter of  $n\_estimators$  was adjusted, and the best value was 71 for the NNI prediction, and 120 for the yield prediction. For the LGMB model, the parameters of 'max\_depth' and 'n\_estimators' were adjusted, the best value was 3590 for NNI prediction, and 4790 for yield prediction.

#### 2.5. Recommendation for N Fertilizer and Density for Different Types of Rice Varieties

The scenario analysis based on the ensemble learning models (RFR and LGMB) were performed to determine the optimum N fertilizer and density for different rice as shown in Figure 1. From historical collected data, the best model was selected with three stages datasets from 2014 to 2021, combined with population structure (leaf area index, SPAD, extinction coefficient), plant shape characteristics (angle of flag leaf, canopy width, plant height), environment, and nutrient utilization. The different type varieties of rice dataset (the large-spike type, intermediate type, and panicle number type) was used to simulate optimum density and N amount based on the established ensemble learning models.



**Figure 1.** The flow chart of recommendation for N fertilizer and density for different types of rice varieties based on ensemble learning model. Note: Population structure is leaf area index, SPAD, and extinction coefficient; Plant shape characteristics is angle of flag leaf, canopy width, and plant height; Environment and nutrition utilization is organic matter, total nitrogen, total phosphorus, slowly available K, available N, available P, available K, value of pH, effective accumulated temperature, active accumulated temperature, and solar energy effective radiation.

2.6. Statistical Analysis

This study collected a total of 295 observations for rice yield and 299 observations for rice NNI. Two independent datasets were generated by utilizing data collected from different sites and years. Specifically, data gathered from Site 1 in 2014, Site 2 in 2017, and Site 3 in 2019 constituted the testing dataset of the model, while the remaining site-year data comprised the training dataset of the model, which were described in Tables 2 and 4.

**Table 4.** The descriptive statistics for the nitrogen nutrition index (NNI) and grain yield across sites and years within the training and the testing datasets.

Dataset	Yield (kg/ha)						NNI					
	n	Mean	SD	Min	Max	CV	n	Mean	SD	Min	Max	CV
Training	236	10,112.8	3624.5	2170.2	18,774.5	36%	239	0.91	0.18	0.47	1.37	20%
Test	59	10,910.7	3896.4	2445.7	18,514	36%	60	0.92	0.20	0.52	1.40	22%

The training dataset was used to build a regression model to predict the NNI and yield, and all models adopted 10-fold cross validation and grid search to adjust optimum parameters. This method divided the dataset into 10 subsets, with one subset used as the validation set in each iteration while the remaining subsets were utilized as the training set. The process was repeated 10 times, selecting a different validation set each time, and then averaging the resulting model performance evaluation metrics to obtain the final assessment. Cross validation enabled fuller use of the dataset and reduced bias introduced by different data divisions [42]. The built regression model was evaluated using the test dataset. Calculate the coefficient of determination ( $R^2$ ), root mean square error (RMSE), and relative error (RE) for the testing datasets using the python programming software (python 3.8) to evaluate the model. The higher the  $R^2$  value and the lower the RMSE and RE values, the better the model. We used OriginPro (Version 2022, OriginLab Corporation, Northampton, MA, USA) and Mathematica (Wolfram Research, Inc., Mathematica, Version 13.0, Champaign, IL, USA, 2021) to generate the plots.

### 3. Results

#### 3.1. Development of a $N_c$ Dilution Curve for Japonica Rice Cultivars and Verification of the $N_c$ Dilution Curve

Plant  $N_c$  was diluted with increase of shoot biomass of plant, with critical N curves fitted following the  $N_c$ -biomass mathematical equations (Figure 2). The equations for the  $N_c$  of japonica rice varieties were shown in Table 5, the coefficients of determination were  $>0.9$  for each cultivars (except DN426). In order to further analyze the significant differences among the 10 rice varieties, the power function model was first linearized to obtain the linearized equations for different varieties (Table 5). Analysis of covariance (ANCOVA) at the 95% confidence interval was used to define the significance of  $N_c$  dilution curves of 10 different rice varieties. The F values and Significance of slope and intercept in Table 6 are statistical difference between different types varieties of rice. To ensure the accuracy of NNI estimation, this study used the  $N_c$ -biomass mathematical equation constructed separately for each variety of rice to calculate the NNI index (Table 5). The predicted  $N_c$  was compared with the measured  $N_c$  data, the lower values of RMSE and higher values of  $R^2$  indicated good stability between the predicted and observed  $N_c$  (Table 7).

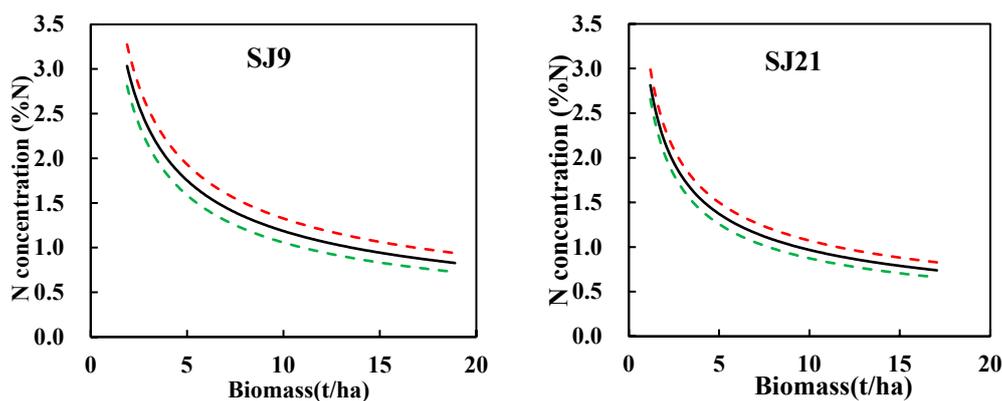
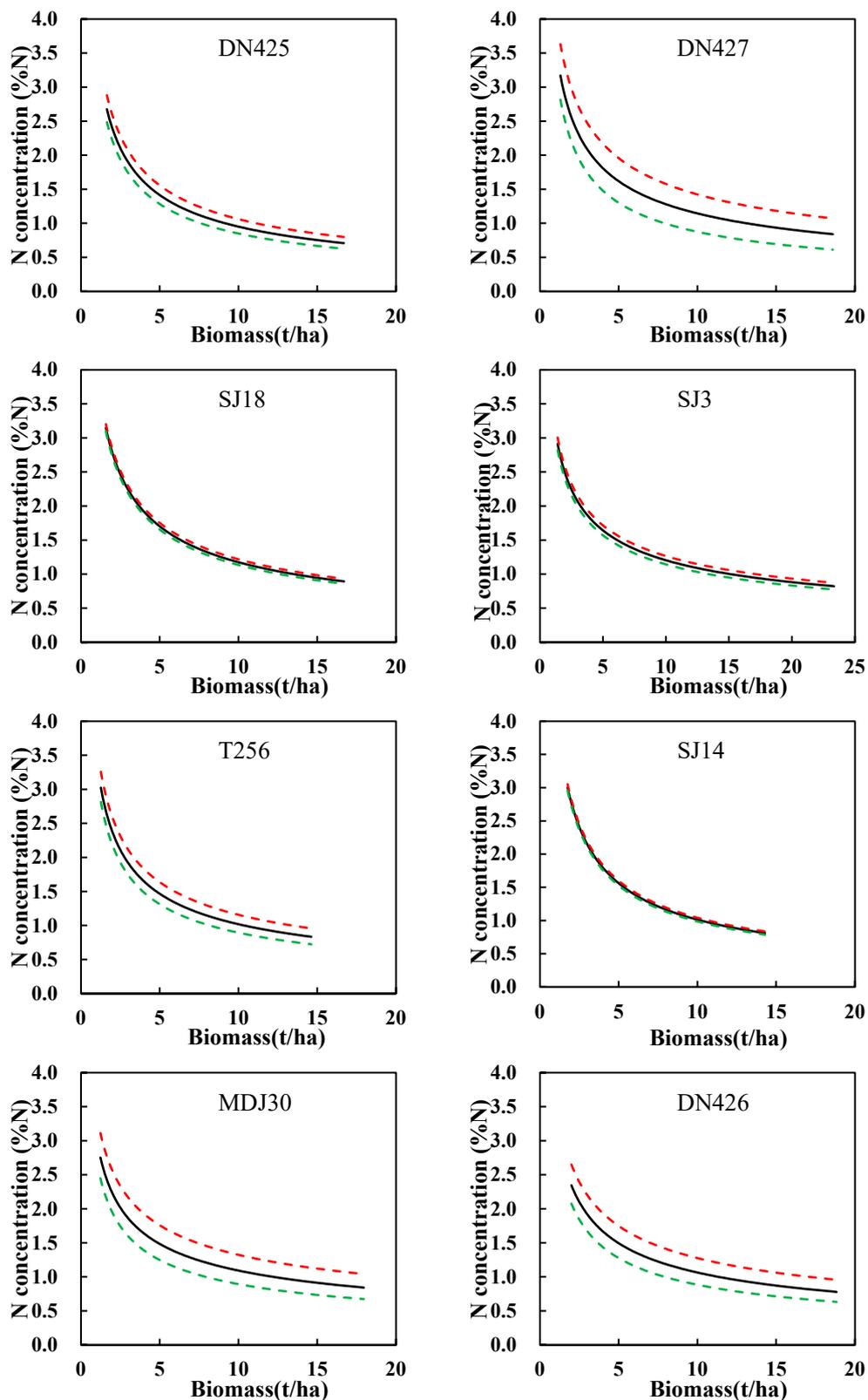


Figure 2. Cont.



**Figure 2.** Critical nitrogen ( $N_c$ ) dilution curve for japonica rice. The solid lines are the  $N_c$  dilution curves describing the relationships between the  $N_c$  and biomass; the dotted lines indicate the 95% confidence band, the red dotted lines indicate indicate 95% confidence interval upper limit and the green dotted lines indicate lower limit.

**Table 5.**  $N_c$  regression equations for different rice varieties.

<i>Japonica</i> Rice Varieties	The Equations for $N_c$	Coefficients of Determination	Power Exponential Function for Linearization Formula
SJ9	$N_c = 4.34W^{-0.563}$	0.9970	$\ln N_c = 1.47 - 0.563 \ln W$
SJ21	$N_c = 3.02W^{-0.503}$	0.9895	$\ln N_c = 1.10 - 0.503 \ln W$
DN425	$N_c = 3.56W^{-0.574}$	0.9932	$\ln N_c = 1.27 - 0.574 \ln W$
DN427	$N_c = 3.66W^{-0.514}$	0.9819	$\ln N_c = 1.30 - 0.514 \ln W$
SJ18	$N_c = 4.03W^{-0.535}$	0.9589	$\ln N_c = 1.39 - 0.535 \ln W$
SJ3	$N_c = 3.39W^{-0.449}$	0.9846	$\ln N_c = 1.22 - 0.449 \ln W$
T256	$N_c = 3.42W^{-0.526}$	0.9887	$\ln N_c = 1.23 - 0.526 \ln W$
SJ14	$N_c = 4.27W^{-0.624}$	0.9671	$\ln N_c = 1.45 - 0.624 \ln W$
MDJ30	$N_c = 3.03W^{-0.443}$	0.9203	$\ln N_c = 1.11 - 0.443 \ln W$
DN426	$N_c = 3.29W^{-0.491}$	0.8493	$\ln N_c = 1.19 - 0.491 \ln W$

**Table 6.** Validation  $N_c$  models for different types of rice varieties.

	Large-Sized Panicle	Small-Sized Panicle	Mid-Sized Panicle	Different Types of Rice
F value	DN425 and SJ14	DN427 and NDJ30	DN426, SJ9, SJ18, SJ21, SJ3 and T256	large-sized panicle, small-sized panicle and mid-sized panicle
Slope	43.613 **	6.895 *	35.011 **	27.031 **
Intercept	62.128 **	10.219 *	39.848 **	24.942 **

Note: \* Significant at  $p < 0.05$ ; \*\* Significant at  $p < 0.01$ .

**Table 7.** Validation of the critical  $N$  dilution curve for different types of rice varieties.

<i>Japonica</i> Rice Varieties	$R^2$	RMSE	RE (%)
SJ9	0.9942	0.0699	2.06
SJ21	0.9891	0.0888	5.34
DN425	0.9921	0.0746	3.11
DN427	0.9809	0.1384	6.13
SJ18	0.9834	0.1254	9.74
SJ3	0.9889	0.0896	5.71
T256	0.9899	0.0957	5.11
SJ14	0.9859	0.1103	8.70
MDJ30	0.9129	0.2088	12.69
DN426	0.8776	0.2175	16.77

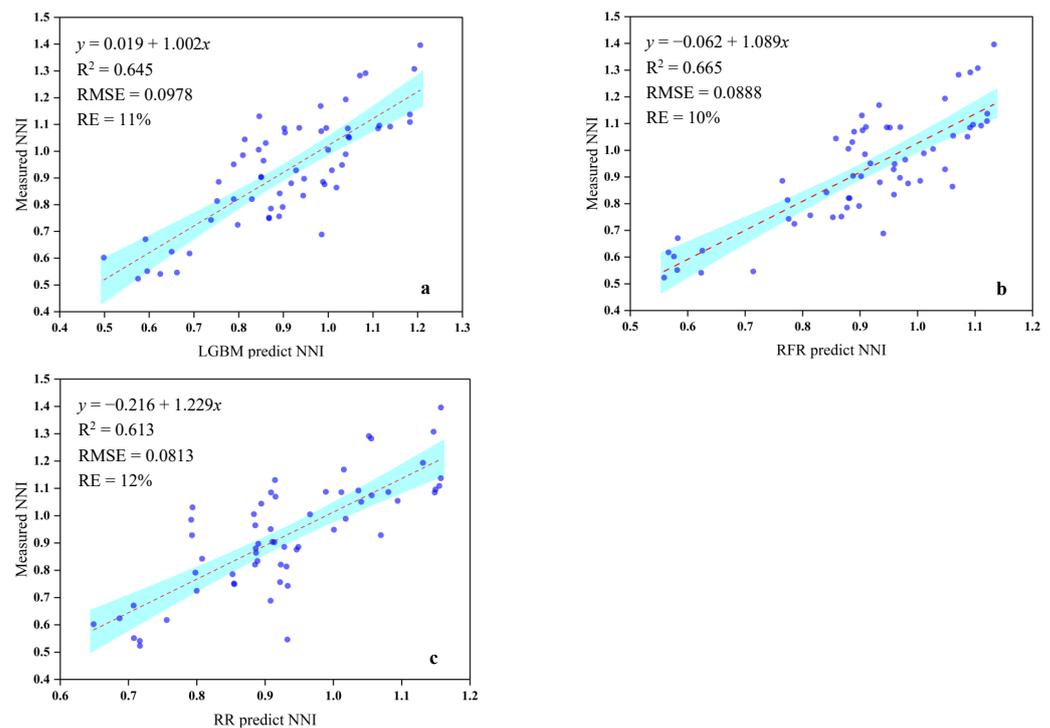
### 3.2. Rice NNI and Grain Yield Variability

A total of 295 rice NNI and 299 yield data points were collected in the experiment. The rice NNI ranged from 0.47 to 1.37 in the training dataset, with a standard deviation of 0.18 and a CV of 20%; and from 0.52 to 1.40 in the test dataset, with a standard deviation of 0.20 and a CV of 22%. The rice yield ranged from 2170.2 to 18,774.5 kg/ha in the training dataset, with a standard deviation of 3624.5 kg/ha and a CV of 36%, and from 2445.7 to 18,514 kg/ha in the test dataset, with a standard deviation of 3896.4 kg/ha and a CV of 36% (Table 4).

### 3.3. The Prediction Model of the Rice NNI

#### 3.3.1. NNI Prediction

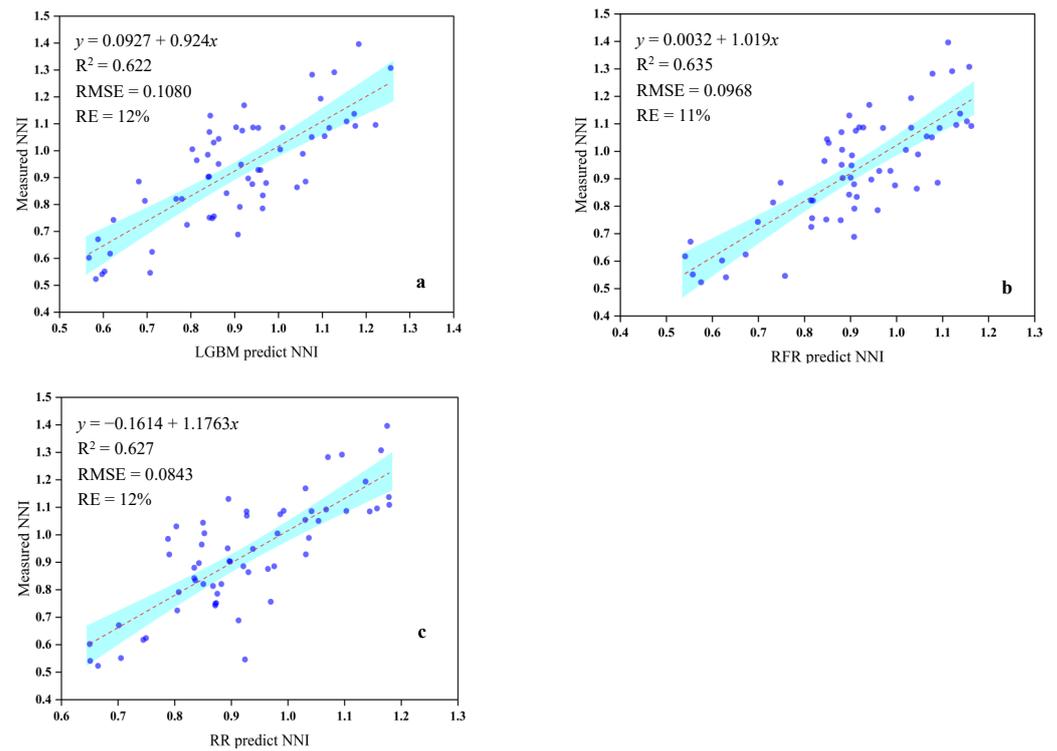
Ridge regression (RR) and ensemble learning methods (RFR and LGBM) were used to predict the NNI for the population structure (leaf area index, SPAD value, extinction coefficient). The test results of the three regression models (RR, RFR, and LGBM) for predicting the rice NNI are shown in Figure 3. Ensemble learning methods performed almost the same (RFR:  $R^2 = 0.665$ , RMSE = 0.0888, RE = 10%; LGBM:  $R^2 = 0.645$ , RMSE = 0.0978, RE = 11%) or better than the RR model ( $R^2 = 0.613$ , RMSE = 0.0813, RE = 12%).



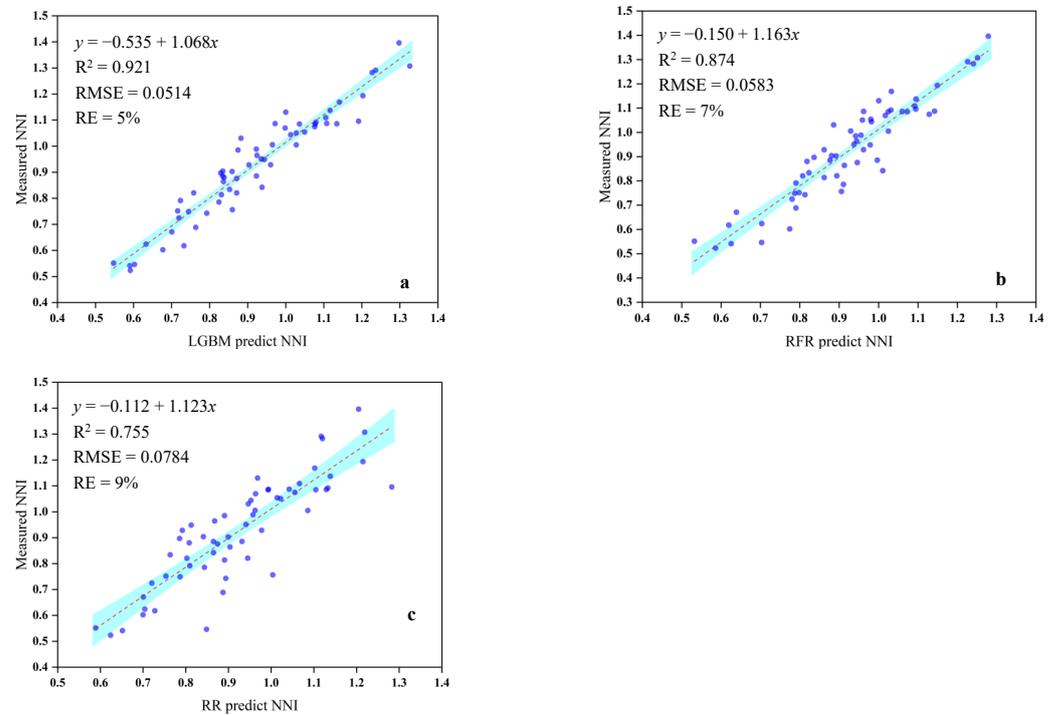
**Figure 3.** The performance of light gradient boosting machine LGBM (a), random forest regression RFR (b) and ridge regression RR (c) to predict nitrogen nutrition index (NNI) for population structure across site-years, N rates and planting densities based on the test dataset.

The RR and ensemble learning methods were used to predict the NNI for the population structure and plant shape characteristics (angle of leaf, canopy width, plant height) (Figure 4). The prediction results of the model after adding the plant shape characteristics are slightly poorer than those obtained when using only a single population structure. RFR performed the best ( $R^2 = 0.635$ , RMSE = 0.0968, RE = 11%), but RR ( $R^2 = 0.627$ , RMSE = 0.0843, RE = 12%) was better than the other ensemble learning method of LGBM ( $R^2 = 0.622$ , RMSE = 0.1080, RE = 12%).

The RR and ensemble learning methods were then used to predict NNI for the population structure, plant shape characteristics and environment and nutrient utilization index (Figure 5). The prediction results of the models were the best after adding the environment and nutrient utilization index. The ensemble models were better than RR ( $R^2 = 0.755$ , RMSE = 0.0784, RE = 9%); specifically, the ensemble model LGBM ( $R^2 = 0.921$ , RMSE = 0.0514, RE = 5%) performed better than RFR ( $R^2 = 0.874$ , RMSE = 0.0583, RE = 7%) (Figure 5).



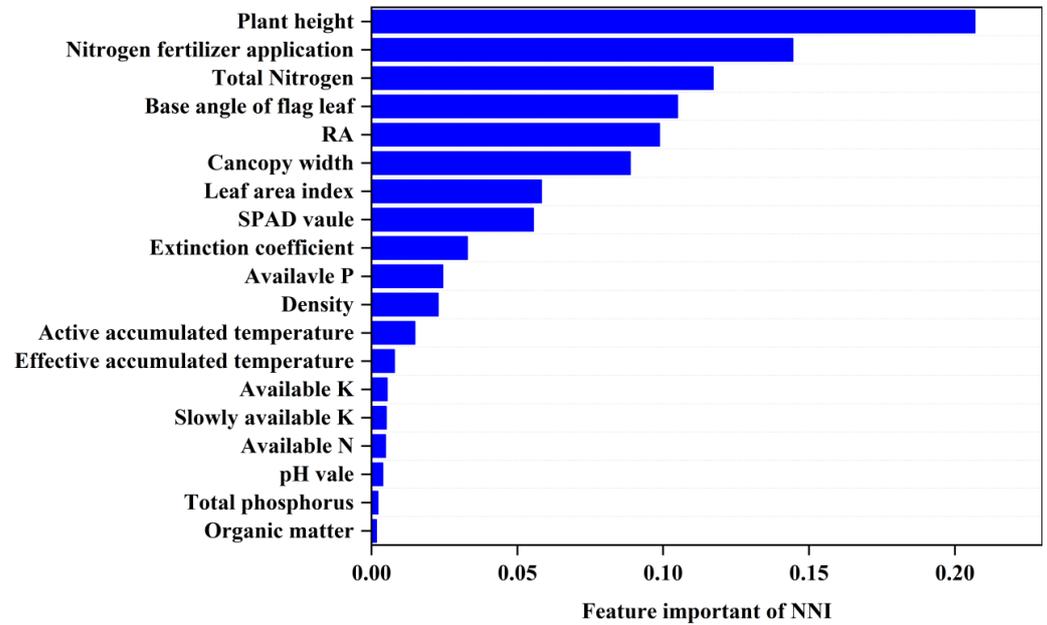
**Figure 4.** The performance of light gradient boosting machine LGBM (a), random forest regression RFR (b) and ridge regression RR (c) to predict nitrogen nutrition index (NNI) for population structure and plant shape characteristics across site-years, N rates and planting densities based on the test dataset.



**Figure 5.** The performance of light gradient boosting machine LGBM (a), random forest regression RFR (b) and ridge regression RR (c) to predict nitrogen nutrition index (NNI) for population structure, plant shape characteristic and environment and nutrient utilization index across site-years, N rates and planting densities based on the test dataset.

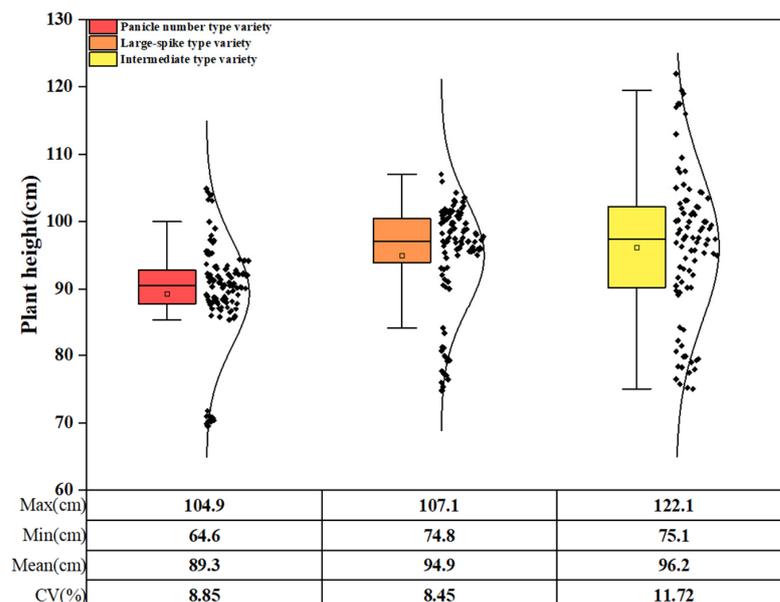
### 3.3.2. Effects of Population Structure and Plant Shape Characteristics on NNI

For NNI prediction, the feature importance in ensemble learning methods is shown in Figure 6 based on the Gini index. The top three features were the plant height (0.207), N fertilizer application (0.144), and total N of soil (0.117).



**Figure 6.** Gini coefficients of input variables for ensemble learning model built on the training dataset to predict nitrogen nutrient index (NNI) across site-years, N rates, and planting densities.

The variation range of the plant height of different types of japonica rice was different, among which, the CV of plant height of intermediate variety was the largest (11.72%), followed by the CV of plant height of the panicle number variety (8.85%), and the CV of plant height of large panicle variety was the smallest (8.45%). The average plant height of the intermediate variety was the largest (96.2 cm), followed by the average plant height of the panicle number variety (94.9 cm), and the average plant height of the panicle grain variety was the smallest (89.3 cm) (Figure 7).



**Figure 7.** Box plot and histogram of plant height of japonica rice with different types.

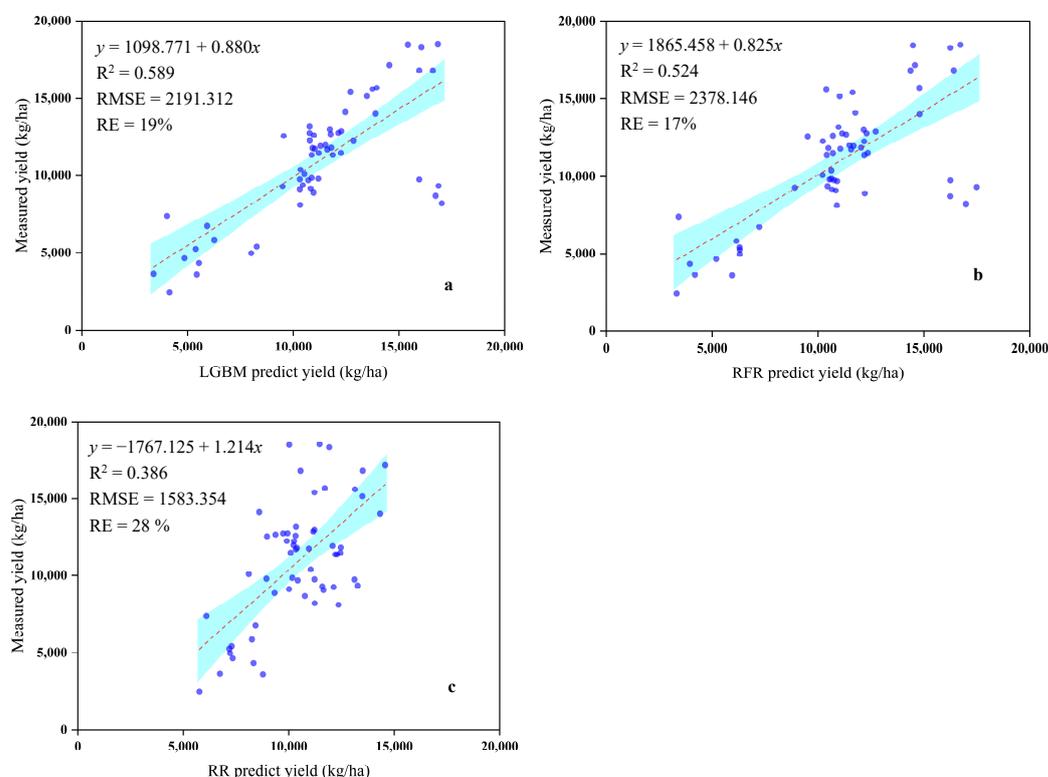
### 3.4. The Prediction Model of Rice Yield

#### 3.4.1. Yield Prediction

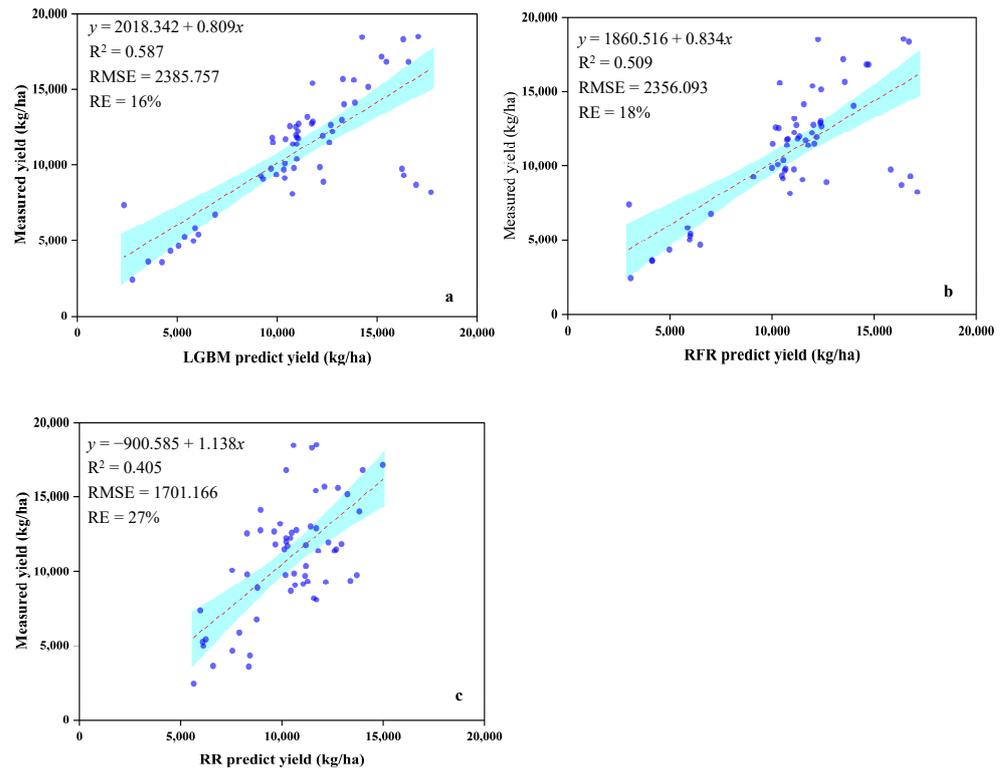
Ridge regression (RR) and ensemble learning methods (RFR and LGBM) were used to predict the yield for the population structure (leaf area index, SPAD value, extinction coefficient). The test results of three regression models (RR, RFR, and LGBM) for predicting the rice yield are shown in Figure 8. The LGBM model ( $R^2 = 0.589$ , RMSE = 2191.312, RE = 19%) performed better than RFR model ( $R^2 = 0.524$ , RMSE = 2378.146, RE = 17%), both ensemble learning models performed better than the RR model ( $R^2 = 0.386$ , RMSE = 1583.354, RE = 28%) (Figure 8).

The RR and ensemble learning methods were used to predict the yield for the population structure and plant shape characteristic angle of leaf, canopy width, plant height) (Figure 9). The ensemble learning methods performed better than RR method. The prediction results of the ensemble learning methods after adding the plant type characteristic are slightly poorer than those obtained when using only a single population structure. The LGBM model performed the best ( $R^2 = 0.587$ , RMSE = 2385.757, RE = 16%), and the RFR performed ( $R^2 = 0.509$ , RMSE = 2356.093, RE = 18%) was better than the RR model ( $R^2 = 0.405$ , RMSE = 1701.166, RE = 27%). The prediction results of the RR method after adding the plant type characteristic are slightly better than that of using only a single population structure.

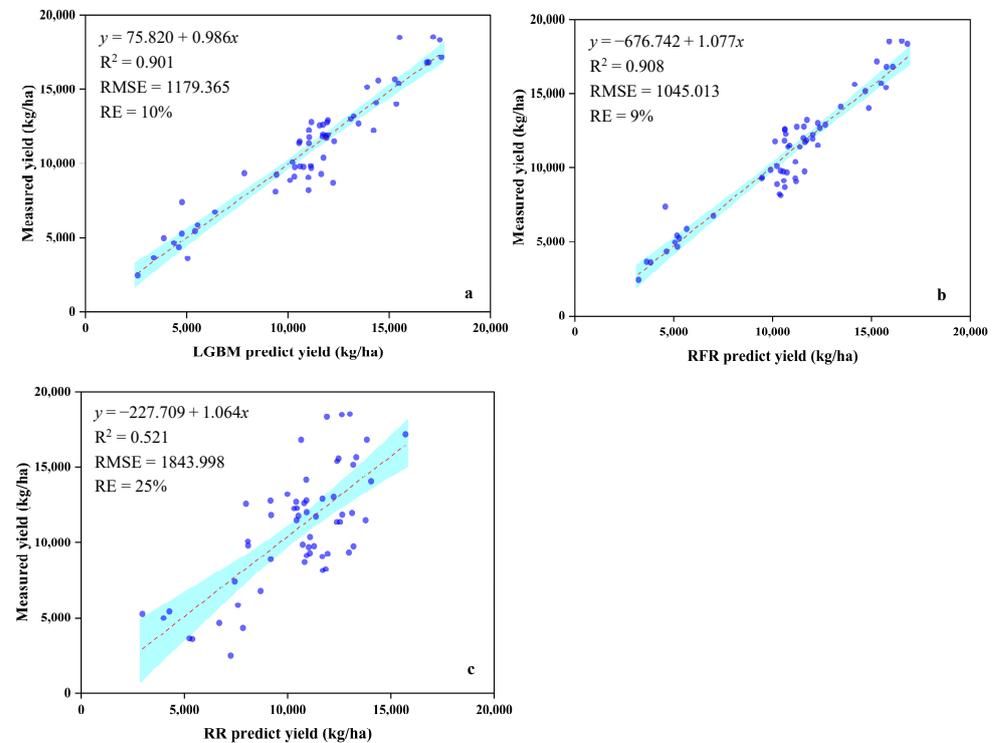
The RR and ensemble learning methods were then used to predict the yield for the population structure, plant shape characteristic and environment and nutrient utilization index (Figure 10). The prediction results of the models are the best after adding environment and nutrient utilization index. The ensemble models were better than the RR models ( $R^2 = 0.521$ , RMSE = 1843.998, RE = 25%); specifically, the ensemble model RFR ( $R^2 = 0.908$ , RMSE = 1045.013, RE = 9%) performed better than the LGBM ( $R^2 = 0.901$ , RMSE = 1179.365, RE = 10%).



**Figure 8.** The performance of light gradient boosting machine LGBM (a), random forest regression RFR (b) and ridge regression RR (c) to predict yield for population structure across site-years, N rates and planting densities based on the test dataset.



**Figure 9.** The performance of light gradient boosting machine LGBM (a), random forest regression RFR (b) and ridge regression RR (c) to predict yield for population structure and plant shape characteristics across site-years, N rates and planting densities based on the test dataset.



**Figure 10.** The performance of light gradient boosting machine LGBM (a), random forest regression RFR (b) and ridge regression RR (c) to predict yield for population structure, plant shape characteristic and environment and nutrient utilization index across site-years, N rates and planting densities based on the test dataset.

### 3.4.2. Effects of Population Structure and Plant Shape Characteristics on Yield

For yield prediction, the feature importance in ensemble learning methods is shown in Figure 11 based on the Gini index. The top three features were the N fertilizer application (0.399), density (0.151) and leaf area index (0.073). The variation range of the leaf area index of different types of *japonica* rice was different (Figure 12), among which, the CV of the leaf area index of the large panicle variety was the largest (30.22%), followed by the CV of the leaf area index of the panicle number variety (19.68%), and the CV of the intermediate variety was the smallest (19.17%). The average leaf area index of the panicle number variety was the largest (6.49), followed by the average leaf area index of the intermediate variety (6.17), and the average leaf area index of the large panicle variety was the smallest (5.48).

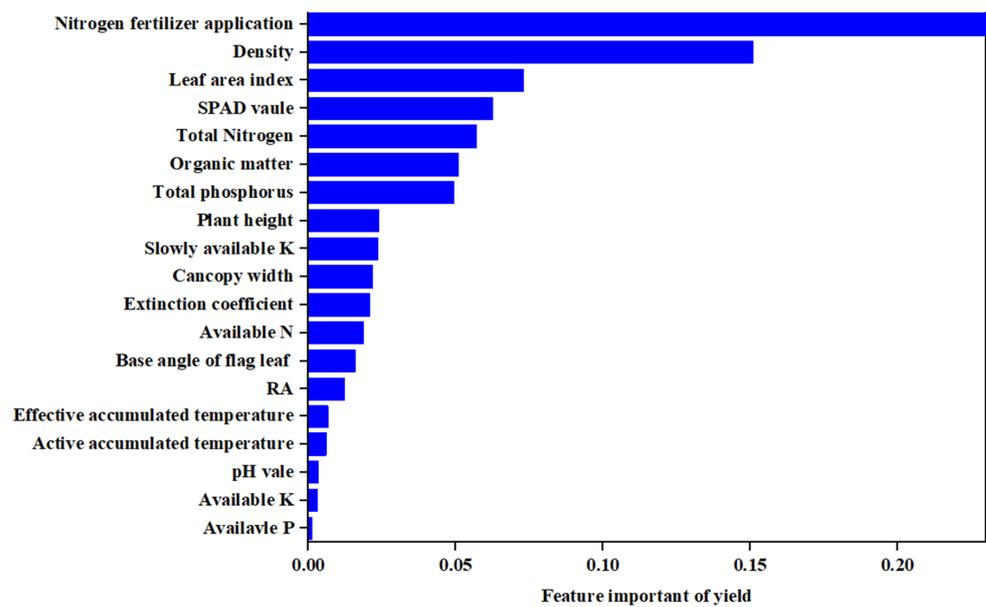


Figure 11. Gini coefficients of input variables for ensemble learning model built on the training dataset to predict yield across site-years, N rates, and planting densities.

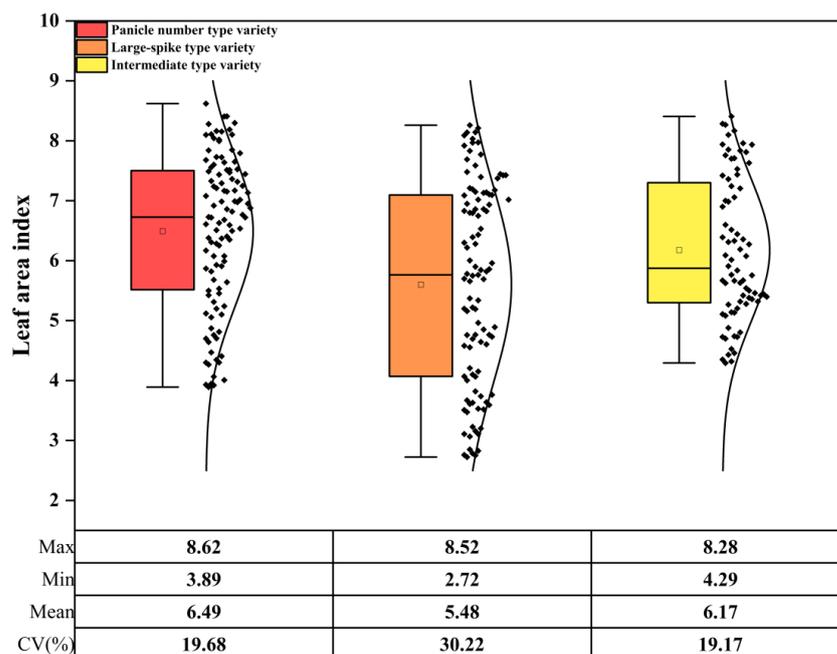


Figure 12. Box plot and histogram of leaf area index of *japonica* rice with different types.

### 3.5. Evaluating Different NNI and Yield Diagnostic Models

While environmental and nutritional factors were added, the prediction performance of the model was significantly improved, in either the linear model or the ensemble learning model. The prediction results of the three models are shown in Table 8.

**Table 8.** The performance of ridge regression, random forest regression and random forest regression models for predicting NNI and grain yield (kg/ha) combining soil, weather, and management variables with vegetation indices across site-years, N rates and planting densities in the training dataset.

Model	Ridge Regression			Random Forest Regression			Light Gradient Boosting Machine		
	R <sup>2</sup>	RMSE	RE (%)	R <sup>2</sup>	RMSE	RE (%)	R <sup>2</sup>	RMSE	RE (%)
NNI	0.755	0.0784	9	0.874	0.0583	7	0.921	0.0514	5
Yield	0.521	1843.998	25	0.908	1045.013	9	0.901	1179.365	10

Although the LGBM model (R<sup>2</sup> = 0.901, RMSE = 1179.365, RE = 10%) was slightly inferior to the RFR model (R<sup>2</sup> = 0.908, RMSE = 1045.013, RE = 9%) in evaluating the yield, the difference was small. The evaluation results of the LGBM model are better than those of the RFR model (R<sup>2</sup> = 0.874, RMSE = 0.0583, RE = 7%) when evaluating the NNI.

### 3.6. Coupling Effect of Density and Fertilizer on Yield and NNI of Rice

#### 3.6.1. Yield of Rice

To pursue high-yield, high-efficiency, high-quality, and sustainable production of rice, this study took yield as the evaluation index to comprehensively evaluate the coupling effect of density and N on rice. Density and N amount were independent variables, and yield was the response variable. Using the prediction results of yield in the third stage of LGBM model (R<sup>2</sup> = 0.901, RMSE = 1179.365, RE = 10%), based on the least squares method, the experimental data were analyzed using Mathematica 13.0 software, and a binary quadratic regression equation was established to calculate the density and N required to maximize the above parameters (Table 9). The optimal system of a single index was solved, and the density and N amount corresponding to the optimal solution of each index in the confidence interval were obtained. The results show that the effects of density and N input on the dependent variable were significant ( $p < 0.01$ ) (Table 9). The density and N amount corresponding to the maximum value of yield of each variety are shown in Table 9.

**Table 9.** Regression equation between the input of density, nitrogen amount and yield in different types of rice varieties. The maximum yield match with the amount of density and nitrogen in different types of rice varieties.

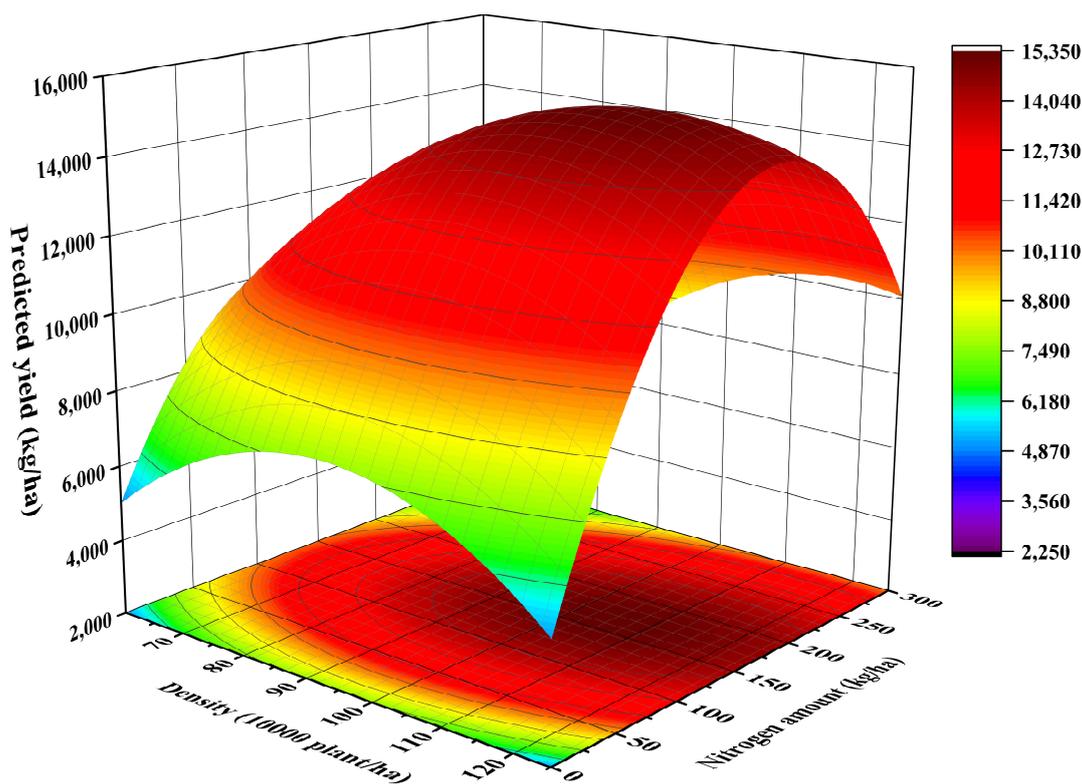
Response Variable Y	Regression Equation	Y Max (kg/ha)	D (10,000 Plant/ha)	N (kg/ha)	R <sup>2</sup>
Yield/Y <sub>1</sub>	$Y_1 = 466.891D + 55.305N - 2.518D^2 - 0.294N^2 + 0.397DN - 13,947.457$	15,314.6	105.768	165.301	0.75
Yield/Y <sub>2</sub>	$Y_2 = 393.942D + 94.142N - 2.193D^2 - 0.246N^2 - 0.200DN - 11,901.280$	11,803.4	82.630	157.826	0.88
Yield/Y <sub>3</sub>	$Y_3 = 89.028D + 53.710N - 0.491D^2 - 0.180N^2 - 0.073DN + 3218.703$	10,392.9	80.815	133.192	0.69

Note: D and N mean density and nitrogen amount, respectively. Y<sub>1</sub> represents the yield of large-spike type variety; Y<sub>2</sub> represents the yield of intermediate type variety; Y<sub>3</sub> represents the yield of panicle number type variety.

As shown in Table 9, the optimum N amount and density were different for different varieties. In the large-spike variety of rice, the maximum yield of 15,314.6 kg/ha was obtained when planting densities of  $105.768 \times 10^4$  plant/ha and 165.301 kg/ha of N were applied. In the intermediate variety of rice, the maximum yield of 11,803.4 kg/ha was obtained when planting densities of  $82.630 \times 10^4$  plant/ha and 157.826 kg/ha of N were applied. In the panicle number variety of rice, the maximum yield of 10,392.9 kg/ha was

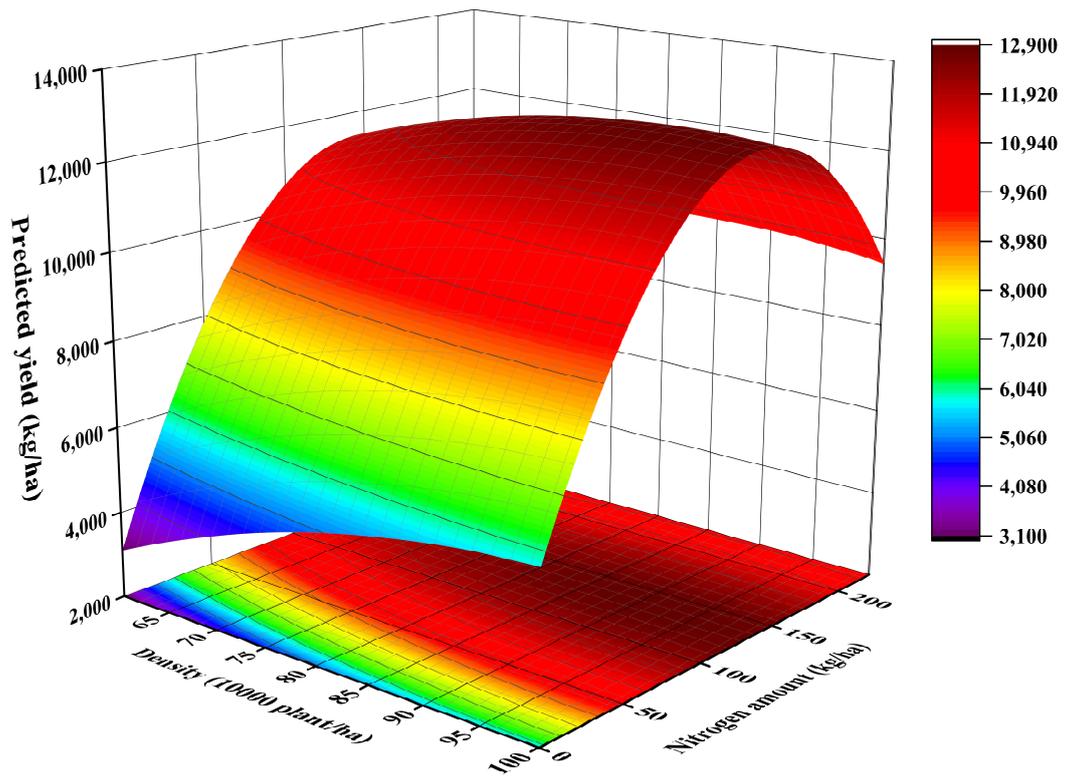
obtained when planting densities of  $80.815 \times 10^4$  plant/ha and 133.192 kg/ha of N were applied.

Because the transplanting density of the rice transplanter was adjusted in units of inches in actual production, the density value calculated in the model could not meet the actual needs of rice transplanting production. Thus, it was necessary to further analyze the density and N amount combination to obtain the best indicators. The coupling effects of density and N amount on the yield in different types of rice varieties exhibited a downward convex shape. Therefore, with the combination of the optimal planting density corresponding to each variety in Table 9 and the actual production situation, the large-spike variety of rice was obtained with the density of 1,001,000 plant/ha (9 inches  $\times$  3 inches), the intermediate variety of rice and the panicle number variety of rice were obtained with the density of 750,800 plant/ha (9 inches  $\times$  4 inches). According to the actual planting density corresponding to different varieties, the yield of different rice varieties and acceptable regions was analyzed, it could be simultaneously obtained within the 99% acceptable range, and the yield ranges were similar (Figure 13). The optimal yield of the large-spike variety of rice was obtained with the density and N amount of  $100.1 \times 10^4$  plant/ha (9 inches  $\times$  3 inches) and 161.05–161.90 kg ha<sup>-1</sup>, respectively. The optimal yield of the intermediate variety of rice was obtained with the density and N amount of  $75.08 \times 10^4$  plant/ha (9 inches  $\times$  4 inches) and 159.52–162.33 kg ha<sup>-1</sup>, respectively. The optimal yield of the panicle number variety of rice was obtained with the density and N amount of  $75.08 \times 10^4$  plant/ha (9 inches  $\times$  4 inches) and 133.47–135.254 kg ha<sup>-1</sup>, respectively.

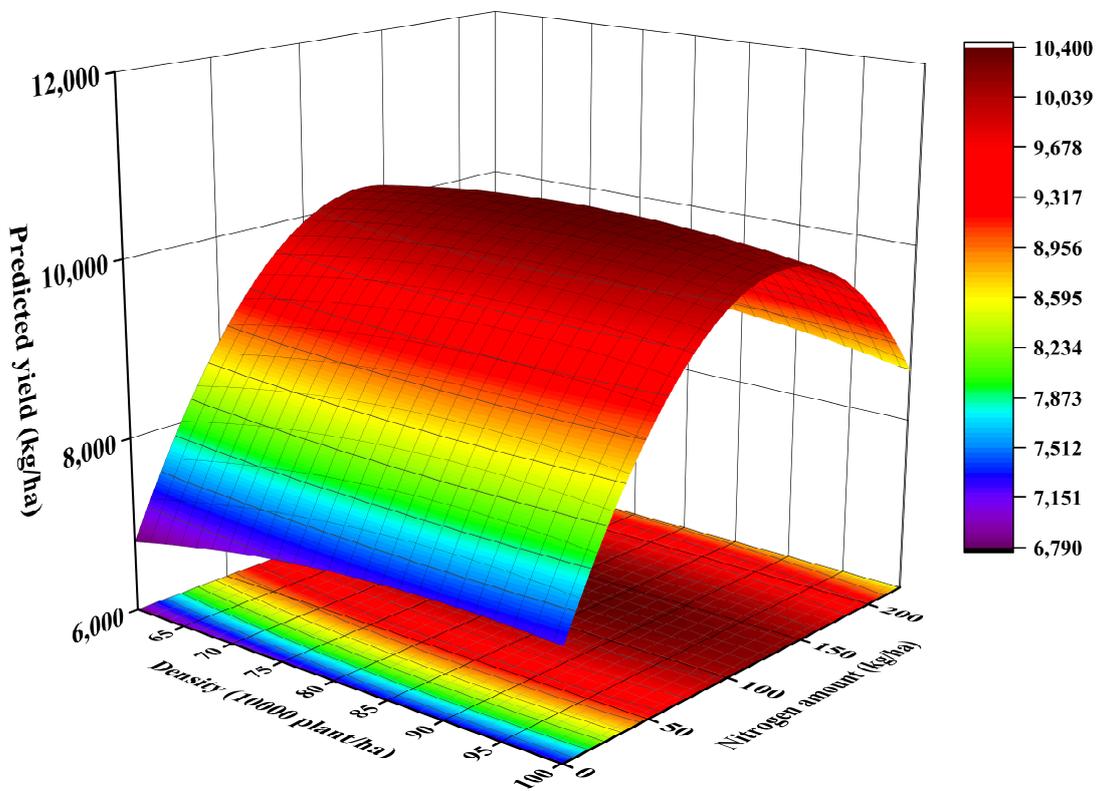


(A)

Figure 13. Cont.



(B)



(C)

**Figure 13.** Relationships between density and nitrogen amount with yield of rice. (A) represents the yield of large-spike type variety, (B) represents the yield of intermediate type variety; (C) represents the yield of panicle number type variety.

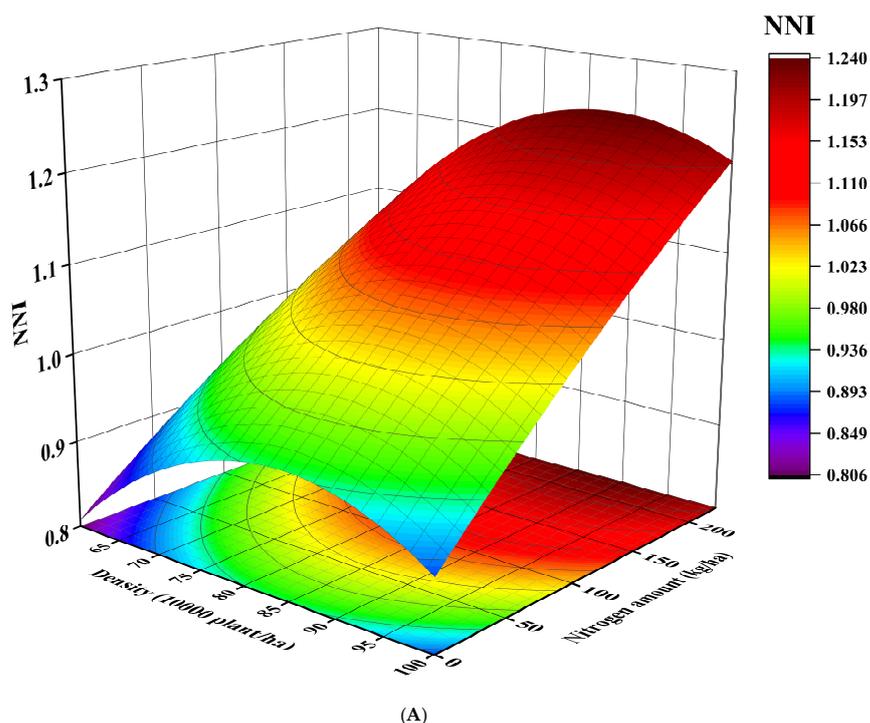
### 3.6.2. NNI of Rice

An NNI of around 1 indicates that the plant N nutrition is sufficient. Using the prediction results of NNI in the third stage of LGBM model (NNI:  $R^2 = 0.921$ , RMSE = 0.0514, RE = 5%), we did not calculate the maximum value of the NNI according to the prediction model, but set the optimal value of NNI to 1. Moreover, three rice varieties with planting densities between  $100.1 \times 10^4$  plant/ha (9 inches  $\times$  3 inches) to  $75.08 \times 10^4$  plant/ha (9 inches  $\times$  4 inches) (Table 10). Then, according to the model, the optimum N fertilizer application rate corresponding to each variety was obtained (Table 10). According to the actual planting density corresponding to different varieties, the NNI of different types of rice varieties and acceptable regions was analyzed, it could be simultaneously obtained within the 99% acceptable range, and the ranges were similar (Figure 14). The optimal yield of the large-spike variety of rice was obtained with the density and N amount of  $100.1 \times 10^4$  plant/ha (9 inches  $\times$  3 inches) and  $53.856\text{--}74.155$  kg ha<sup>-1</sup>, respectively. The optimal yield of the intermediate variety of rice was obtained with the density and N amount of  $75.08 \times 10^4$  plant/ha (9 inches  $\times$  4 inches) and  $161.406\text{--}194.003$  kg ha<sup>-1</sup>, respectively. The optimal yield of the panicle number variety of rice was obtained with the density and N amount of  $75.08 \times 10^4$  plant/ha (9 inches  $\times$  4 inches) and  $121.691\text{--}193.900$  kg ha<sup>-1</sup>, respectively.

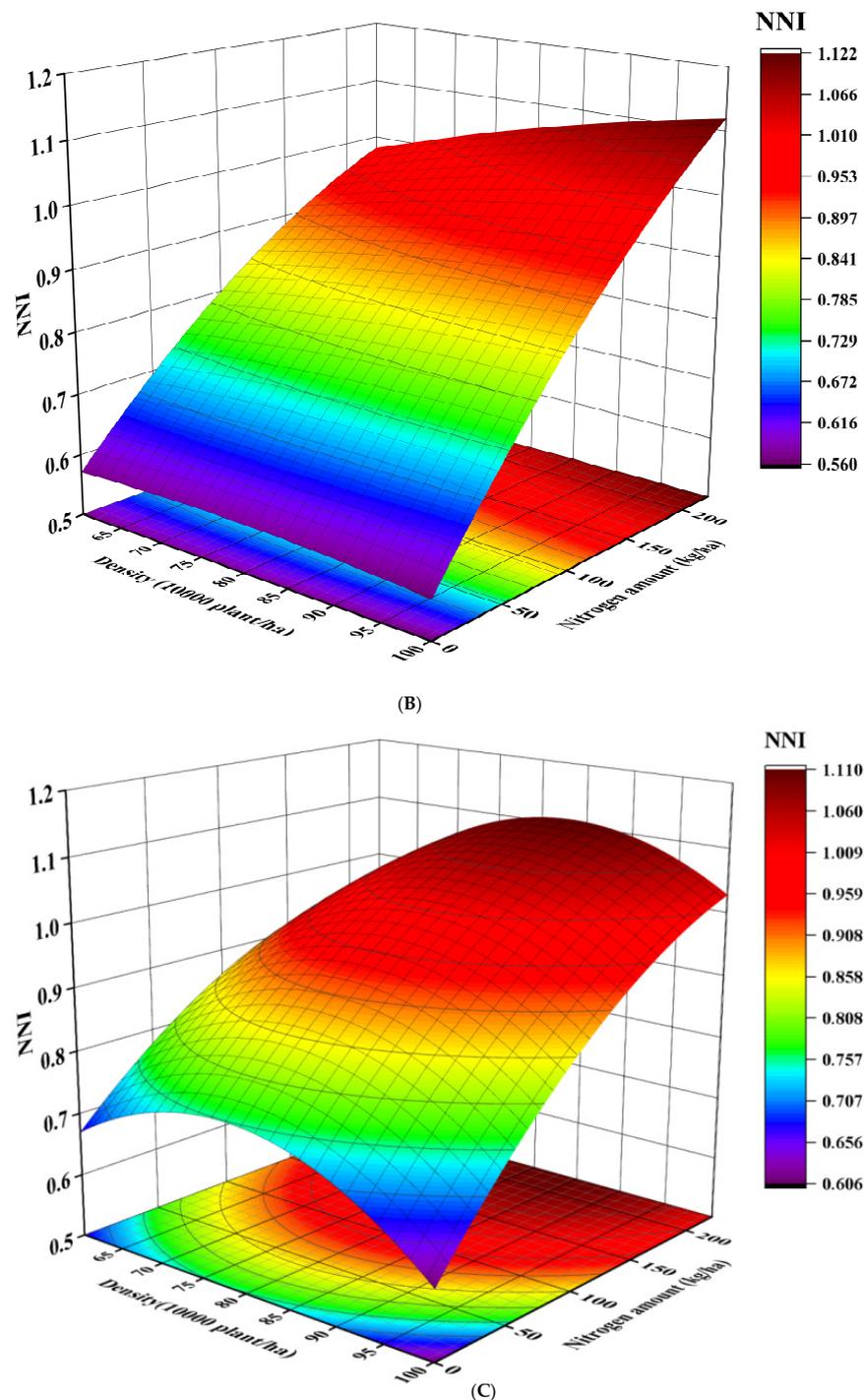
**Table 10.** Regression equation between the input of density, nitrogen amount and NNI in different types of rice varieties. The optimal value of NNI match with the amount of density and nitrogen in different types of rice varieties.

Response Variable Y	Regression Equation	Y Max	D (10,000 Plant/ha)	N (kg/ha)	R <sup>2</sup>
NNI/Y <sub>4</sub>	$Y_4 = 3.879 \times 10^{-2}D + 8.633 \times 10^{-4}N - 2.309 \times 10^{-4}D^2 - 5.218 \times 10^{-6}N^2 + 7.402 \times 10^{-6}DN - 0.689$	1	$75.08 \leq x \leq 100.1$	$53.856 \leq N \leq 74.155$	0.62
NNI/Y <sub>5</sub>	$Y_5 = 3.800 \times 10^{-3}D + 1.640 \times 10^{-3}N - 2.522 \times 10^{-5}D^2 - 3.663 \times 10^{-6}N^2 + 1.665 \times 10^{-5}DN + 0.434$	1	$75.08 \leq x \leq 100.1$	$161.406 \leq N \leq 194.003$	0.82
NNI/Y <sub>6</sub>	$Y_6 = 4.456 \times 10^{-2}D + 1.520 \times 10^{-3}N - 2.879 \times 10^{-4}D^2 - 5.083 \times 10^{-6}N^2 + 1.490 \times 10^{-5}DN - 0.968$	1	$75.08 \leq x \leq 100.1$	$121.691 \leq N \leq 193.900$	0.74

Note: D and N mean density and nitrogen amount, respectively. Y<sub>4</sub> represents the yield of large-spike type variety; Y<sub>5</sub> represents the yield of intermediate type variety; Y<sub>6</sub> represents the yield of panicle number type variety.



**Figure 14.** Cont.



**Figure 14.** Relationships between density and nitrogen amount with NNI of rice. (A) represents the NNI of large-spike type variety, (B) represents the NNI of intermediate type variety; (C) represents the NNI of panicle number type variety.

We found that the yield and NNI model prediction results are considerably different between the large-panicle varieties and the intermediate varieties. Large-spiked varieties require that the nitrogen concentration of the plant is in a surplus state to obtain a high yield, while intermediate varieties can achieve higher yields when the plants are in a state of N concentration deficiency. Moreover, the panicle number varieties had the same basic trend in the combined yield and NNI model prediction results (Tables 9 and 10). Combining the yield and NNI model prediction results, we finally determined that the optimal yield ( $15,327.99 \text{ kg ha}^{-1}$ ) and NNI (1.12) of the large-spike type variety of rice were

obtained with the density and N amount of  $100.1 \times 10^4$  plant/ha (9 inches  $\times$  3 inches) and  $161.05 \text{ kg ha}^{-1}$ , respectively; the optimal yield ( $11,680.30 \text{ kg ha}^{-1}$ ) and NNI (0.89) of the intermediate type variety of rice were obtained with the density and N amount of  $75.08 \times 10^4$  plant/ha (9 inches  $\times$  4 inches) and  $159.52 \text{ kg ha}^{-1}$ , respectively; and the optimal yield ( $10,376.89 \text{ kg ha}^{-1}$ ) and NNI (1.01) of panicle number type variety of rice were obtained with the density and N amount of  $75.08 \times 10^4$  plant/ha (9 inches  $\times$  4 inches) and  $133.47 \text{ kg ha}^{-1}$ , respectively.

#### 4. Discussion

##### 4.1. Comparison of the Regression Models for Predicting NNI and Grain Yield

The N application amount and transplanting density are important field management factors that affect the rice yield [2,7]. Most of the relevant research on precision agriculture models has focused on the effect of N fertilizer application on crop yield [5,9], while there have been few studies related to precision agriculture models on the effects of N application and density on rice yield [10]. It is a promising approach to determine optimal N rates under different soil and weather conditions [43]. Some studies suggested that there was no difference in the critical N% changes of japonica rice in the same ecological type. Usually, data from different cultivars of the same subtype of rice were combined to improve the applicability of the curve [44,45]. However, the results of this study did not entirely support this conclusion; the F values and significance of slope and intercept of Nc–biomass mathematical equation are statistically different between different rice type varieties (Table 6). Many factors involving genetics, the environment, and management affected critical N concentration dilution curve of rice [44,46]. Determining which N dilution curve to use for data calculation is crucial. To ensure the accuracy of NNI estimation, this study used the Nc–biomass mathematical equation constructed separately for each variety of rice to calculate the NNI index (Table 5). In order to explore optimal N application and density for rice, the parameters of actual environmental conditions (e.g., soil properties and weather conditions) and crop management information in production should be considered when predicting the NNI and yield and N fertilizer and density management.

Multiple linear regression models can integrate multiple factor data for model building and data prediction [47]. The RR model is an empirical approach to the multiple linear regression model, and it can show correlations with the yield and integrate morpho-physiological indexes to determine the yield [48]. In this study, the RR model using the PS data (leaf area index, SPAD value, extinction coefficient) could only explain 61.3% and 38.6% of the NNI and yield variability based on the test dataset, respectively (Figures 3 and 8). This result is similar to that of studies that characterized the influence of initial conditions, crop management, soil management, and N fertilizer management, which were used to predict corn yield under different regions ( $R^2 = 0.350$ ) [49]. Compared with the RR model using only the PS data, the prediction effect of the RR model improved the accuracy of the NNI and yield prediction by using the PS and PSC data (NNI:  $R^2 = 0.627$ ; yield:  $R^2 = 0.405$ ) and by using PS, PSC and ENI (NNI:  $R^2 = 0.755$ ; yield:  $R^2 = 0.521$ ). Previous research reported improved NNI and grain yield prediction accuracy when multiple linear regression models with the remote sensing data, soil data, weather data, and management data were used [5,10]. In this study, the prediction of the NNI by the linear model was significantly better than that of the yield. This may be because the NNI is closely related to the population structure, such as LAI [50,51] and plant height, and the model for predicting the NNI based on LAI has been confirmed by a large number of studies [52,53].

Over the past decade, ML methods have developed rapidly in the field of precision agriculture [5,10,25–29]. ML and deep learning methods can solve significant nonlinear problems using datasets from multiple sources. On the contrary, linear regression models have limitations in analyzing nonlinear relationships, high-order interactions, and non-normal data [54]. At present, most of the PNM model predictions about the NNI and yield level use the RFR model [5,28,29,38], while there have been few related studies on

the prediction of the yield and NNI using the LGBM model that focus on the ensemble learning method. Previous research found that the RFR model was better at predicting the corn NNI and grain yield than linear regression [5]. In this study, the ensemble learning model (RFR and LGBM) was better than the linear model (Figures 3–5 and 8–10), which is consistent with the results of previous studies [5].

In contrast, in the second stage, the NNI and yield prediction effect of the ensemble model (NNI: LGBM model  $R^2 = 0.622$ , RFR model  $R^2 = 0.635$ ; yield: LGBM model  $R^2 = 0.587$ , and RFR model  $R^2 = 0.509$ ) did not improve but decreased compared with that of the first stage (NNI: LGBM model  $R^2 = 0.645$ , RFR model  $R^2 = 0.665$ ; yield: LGBM model  $R^2 = 0.589$ , and RFR model  $R^2 = 0.524$ ) (Figures 3 and 8). This is different from the result of a previous study, which found that the ensemble model outperformed when using multiple types of data [49]. There are two main reasons for this situation. First, the modeling data used by predecessors did not include experimental data such as population plant types [5]; however, the population plant type data and NNI and yield had a complex relationship with the variety [18], fertilizer dosage [55], transplanting density [19], and cultivation environment conditions [10]. Another possible reason is the relatively small dataset size, which may have limited the performance of the ML methods.

Moreover, in the third stage, the NNI and yield prediction effect of the ensemble learning model was significantly improved compared to that of the other two stages, and the  $R^2$  of LGBM and RFR in yield reached 0.901 and 0.908, respectively, while the  $R^2$  of LGBM and RFR in NNI reached 0.921 and 0.874, respectively. This is consistent with most research findings [5,12,28].

#### 4.2. N and Density Recommendation Based on the Ensemble Learning Model

Previous studies have demonstrated that a small dataset (<500) can be used to build models for predictive analysis with deep learning [56]. This study aimed to evaluate the predictive performance of ensemble learning model for N and density recommendation in the northeast China under wide range of conditions in three types of rice, for a small dataset (<500). Both N and density are key information for developing methods for on-farm crop management and growth monitoring [56,57]. However, there is very little known about the effect of the small and diverse dataset with three types of rice on the predictive performance of various ensemble learning models such as RFR and LGBM. We developed, tested, and also inter-compared two ensemble learning models for their generalizability on independent data.

The determination of the N and density recommendation for different varieties of rice is challenging because of the interactions of crop variety [58], environmental weather conditions [5], soil properties [34], and field management techniques [35]. Most of the data used for the prediction by ML models in the past were obtained by sensors [12,25,28], and owing to the difficulty of traditional data collection, the model prediction of yield and the NNI has not been combined with characteristic value data, such as the population structure or plant shape characteristics. In addition, most crop growth model-based approaches to predict optimal N rate used average N rates across many years or sites, but are not site-specific [59]. Therefore, there is an urgent need to develop a PNM strategy combining plant population structure and plant shape characteristics for sustainable rice production in northeast China.

Yield, N application, and planting density have a quadratic, quadratic plus plateau, or linear plus plateau relationship with the N application rate and are only linear in rare cases [60–63]. As a result, it was difficult to determine the optimum N rate and density that would maximize grain yield and NNI in a linear fashion (RR) (Figures 5 and 10). Therefore, the RR model was not suitable for determining the optimum N rate and density of rice (NNI:  $R^2 = 0.755$ ; yield:  $R^2 = 0.521$ ) (Figures 5 and 10).

At present, most of the PNM model predictions about the NNI and yield level use the RFR model [5,28,29,38]. However, there have been few related studies on the prediction of the yield and NNI using the LGBM model, which focus on the ensemble learning method.

Because the LGBM model is an efficient gradient boosted decision tree, it speeds up the training process and is sometimes more accurate [26]. In this study, based on the accuracy of predictions, an innovative N and density recommendation strategy was developed using the RFR prediction model to simulate yield (Figure 10) and the LGBM prediction model was used to simulate NNI responses to a series of N application and density management scenarios (Figure 5).

#### 4.3. The Application Potential of N and Density Recommendation Model Based on Multiple Regression Analysis

Predecessors have used the multiple regression analysis method to comprehensively evaluate the effects of water and N fertilizers on the yield, agronomic traits, NUE, and water use efficiency of different crops [5]. These studies pointed out that the multiple regression analysis method can accurately quantify the optimal value of the target and meet different requirements rather than comparing trends and simple size [64]. In this study, we further analyzed the relationships between the N rate, density, yield, and the NNI by considering for the maximum values of different types of rice varieties of yield and a suitable NNI based on the analysis of Mathematic 13.0 (Tables 9 and 10; Figures 13 and 14). Because the NNI is a reliable indicator for diagnosing the N status of crops, an NNI of around one indicates that the plant N nutrition is sufficient [65]. Therefore, in this study we set the optimal value of NNI to one. We found that high-density ( $100.1 \times 10^4$  plant/ha; 9 inches  $\times$  3 inches) and high-N rate ( $161.05 \text{ kg ha}^{-1}$ ) cultivation techniques could be selected for large panicle cultivars, and a high yield ( $15,327.99 \text{ kg ha}^{-1}$ ) could be obtained by the slight surplus status of the NNI (1.12) (Tables 9 and 10; Figures 13 and 14). Relevant studies have found that high-N fertilizer ( $220 \text{ kg ha}^{-1}$ ) and low-density levels ( $13.1 \times 10^4$  hills  $\text{ha}^{-1}$ ) can enable low-tillering and large-sized panicle cultivar japonica rice varieties to achieve high yields [19], which is inconsistent with the results of this study. In contrast, the intermediate varieties can obtain a high yield ( $11,680.30 \text{ kg ha}^{-1}$ ) under the conditions of medium-density ( $75.08 \times 10^4$  plant/ha; 9 inches  $\times$  4 inches) and medium-N rate ( $159.52 \text{ kg ha}^{-1}$ ) cultivation techniques, even if the plants are in a state of N deficiency (NNI, 0.89) (Tables 9 and 10; Figures 13 and 14). Some studies have reported that dense planting is a feasible strategy to reduce N input in inbred rice [57], japonica inbred rice [19], and perennial rice [66], which can enable high-yield rice. This is not the same as the results of this study, our findings indicate that the intermediate varieties can achieve high yields under moderate to high-N fertilizer conditions and at lower planting densities. For the panicle number type variety, the N rate ( $133.47 \text{ kg ha}^{-1}$ ) can be appropriately reduced under a condition of medium density ( $75.08 \times 10^4$  plant/ha; 9 inches  $\times$  4 inches), and a high yield ( $10,376.89 \text{ kg ha}^{-1}$ ) can be obtained when the NNI (1.01) is maintained at about one (Tables 9 and 10; Figures 13 and 14). In contrast, other studies have reported that high-tillering rice cultivars should be planted with more seedlings per hill to obtain a higher yield [19].

In summary, according to previous studies and the results of this study, it can be concluded that there is no uniform optimal planting density, owing to the different cultivar characteristics. Therefore, the method based on ML can determine the optimal planting density and fertilization amount for different kinds of varieties, which can effectively exert the characteristics of varieties and achieve the purpose of increasing yield.

Note that this study only covered one soil type, three different cultivar types, and several years of weather data. More annual site-N rate experiments need to be conducted under different soil and weather conditions to train and test the ensemble learning model and thereby more accurately evaluate the N fertilizer and density recommendation strategy based on the ensemble learning method. Moreover, the potential of the ensemble learning model for N and density recommendation strategies can be optimized based on the data acquired from sensors such as Crop Circle ACS 430 and RapidSCAN CS-45 [67,68]. In addition, unmanned aerial vehicle remote sensing or satellite remote sensing can also be used to monitor the N

status of large-scale crops and improve the ensemble learning model so as to guide the N management of japonica rice in cold regions in the Heilongjiang Province [10].

## 5. Conclusions

This study compared the performance of the RR, RFR, and LGBM models for prediction of NNI and yield using PS, PSC, and ENI data and developed an ensemble learning model-based N and D recommendation strategy for different types of rice. The results demonstrate that in the three stages of prediction of NNI and yield, the prediction effect of the ensemble learning model was significantly better than that of the linear model. The prediction effect of the linear model improved gradually with the increase in the number of types of data (NNI:  $R^2 = 0.613, 0.627, 0.755$ ; yield:  $R^2 = 0.386, 0.405, 0.521$ ). In the third stage, the NNI and yield prediction effect of the ensemble learning model was more significantly improved than that of the other two stages; the  $R^2$  of LGBM and RFR in yield reached 0.901 and 0.908, respectively; and the  $R^2$  of LGBM and RFR in NNI reached 0.921 and 0.874, respectively. An innovative N and density recommendation strategy was developed using the RFR prediction model to simulate yield and the LGBM prediction model to simulate NNI responses to a series of N application and density management scenarios. The scenario analysis results show that the optimal yield ( $15,327.99 \text{ kg ha}^{-1}$ ) and NNI (1.12) of the large-spike variety of rice were obtained with a density and N amount of  $100.1 \times 10^4 \text{ plant/ha}$  (9 inches  $\times$  3 inches) and  $161.05 \text{ kg ha}^{-1}$ , respectively; the optimal yield ( $11,680.30 \text{ kg ha}^{-1}$ ) and NNI (0.89) of the intermediate variety of rice were obtained with a density and N amount of  $75.08 \times 10^4 \text{ plant/ha}$  (9 inches  $\times$  4 inches) and  $159.52 \text{ kg ha}^{-1}$ , respectively; and the optimal yield ( $10,376.89 \text{ kg ha}^{-1}$ ) and NNI (1.01) of panicle number variety of rice were obtained with a density and N amount of  $75.08 \times 10^4 \text{ plant/ha}$  (9 inches  $\times$  4 inches) and  $133.47 \text{ kg ha}^{-1}$ , respectively. It is concluded that the ensemble learning model-based N and D recommendation strategy combining crop sensing data with PS, PSC, and ENI information is a promising approach to improve rice N and D management.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy14051028/s1>, Figure S1: The location of the study sites; the blue dot represents the Site 1 location, the red dot represents the Site 2 location, and the pink dot represents the Site 3 location.

**Author Contributions:** Y.J. and Y.Z.: conceptualization, data curation, formal analysis, funding acquisition, investigation, and methodology; H.M.: investigation, methodology, and formal analysis; D.Z.: project administration, supervision, writing—review and editing; W.G.: investigation, resources, and formal analysis; J.W. (Jingguo Wang): methodology, project administration, and supervision; J.W. (Jin Wang): methodology, resources, software, and formal analysis; C.Z.: project administration, visualization, and software. A.L.: project administration, visualization, formal analysis, and software; W.W.: project administration, visualization, and formal analysis; P.X.: investigation, formal analysis, and software; Q.Y., J.W. (Jing Wang) and Z.W.: conceptualization, software, validation, and formal analysis; H.Z.: funding acquisition, supervision, writing—review and editing. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China (No. 32301935), the Heilongjiang Province Applied Technology Research and Development Plan Project (No. GA20B101), the Heilongjiang Province Natural Science Foundation Project (No. LH2020C005), the Postdoctoral Fund to Research Start-up of Heilongjiang Province (No. LBH-Q21077), the Ministry of Education Industry–University Cooperative Education Program (No. 202101001030), and the Agricultural Ecological Resources and Environmental Protection Service Project of the Ministry of Agriculture and Rural Affairs (No. 13220061 and No.13230055).

**Data Availability Statement:** All data are available via an email request to the authors.

**Conflicts of Interest:** Author Jin Wang was employed by the Bei Da Huang Kenfeng Seed Limited Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## References

1. FAO FaAO. *The Future of Food and Agriculture-Trends and Challenges* (Rome: Food and Agriculture Organization); FAO: Rome, Italy, 2017.
2. Hou, W.; Khan, M.R.; Zhang, J.; Lu, J.; Ren, T.; Cong, R.; Li, X. Nitrogen rate and plant density interaction enhances radiation interception, yield and nitrogen use efficiency of mechanically transplanted rice. *Agric. Ecosyst. Environ.* **2019**, *269*, 183–192. [[CrossRef](#)]
3. Song, X.; Meng, X.; Guo, H.; Cheng, Q.; Jing, Y.; Chen, M.; Liu, G.; Wang, B.; Wang, Y.; Li, J. Targeting a gene regulatory element enhances rice grain yield by decoupling panicle number and size. *Nat. Biotechnol.* **2022**, *40*, 1403–1411. [[CrossRef](#)] [[PubMed](#)]
4. Yamano, T.; Arouna, A.; Labarta, R.A.; Huelgas, Z.M.; Mohanty, S. Adoption and impacts of international rice research technologies. *Glob. Food Secur.* **2016**, *8*, 1–8. [[CrossRef](#)]
5. Wang, X.; Miao, Y.; Dong, R.; Zha, H.; Xia, T.; Chen, Z.; Kusnierek, K.; Mi, G.; Sun, H.; Li, M. Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn. *Eur. J. Agron.* **2021**, *123*, 126193. [[CrossRef](#)]
6. Yuan, S.; Linqvist, B.A.; Wilson, L.T.; Cassman, K.G.; Stuart, A.M.; Pede, V.; Miro, B.; Saito, K.; Agustiani, N.; Aristya, V.E. Sustainable intensification for a larger global rice bowl. *Nat. Commun.* **2021**, *12*, 7163. [[CrossRef](#)] [[PubMed](#)]
7. Li, P.; Lu, J.; Wang, Y.; Wang, S.; Hussain, S.; Ren, T.; Cong, R.; Li, X. Nitrogen losses, use efficiency, and productivity of early rice under controlled-release urea. *Agric. Ecosyst. Environ.* **2018**, *251*, 78–87. [[CrossRef](#)]
8. Yost, M.; Kitchen, N.R.; Sudduth, K.A.; Massey, R.; Sadler, E.; Drummond, S.; Volkmann, M. A long-term precision agriculture system sustains grain profitability. *Precis. Agric.* **2019**, *20*, 1177–1198. [[CrossRef](#)]
9. Wang, X.; Miao, Y.; Dong, R.; Chen, Z.; Guan, Y.; Yue, X.; Fang, Z.; Mulla, D.J. Developing active canopy sensor-based precision nitrogen management strategies for maize in Northeast China. *Sustainability* **2019**, *11*, 706. [[CrossRef](#)]
10. Zha, H.; Miao, Y.; Wang, T.; Li, Y.; Zhang, J.; Sun, W.; Feng, Z.; Kusnierek, K. Improving unmanned aerial vehicle remote sensing-based rice nitrogen nutrition index prediction with machine learning. *Remote Sens.* **2020**, *12*, 215. [[CrossRef](#)]
11. Zhou, B.; Sun, X.; Wang, D.; Ding, Z.; Li, C.; Ma, W.; Zhao, M. Integrated agronomic practice increases maize grain yield and nitrogen use efficiency under various soil fertility conditions. *Crop J.* **2019**, *7*, 527–538. [[CrossRef](#)]
12. Lu, J.; Wang, H.; Miao, Y.; Zhao, L.; Zhao, G.; Cao, Q.; Kusnierek, K. Developing an Active Canopy Sensor-Based Integrated Precision Rice Management System for Improving Grain Yield and Quality, Nitrogen Use Efficiency, and Lodging Resistance. *Remote Sens.* **2022**, *14*, 2440. [[CrossRef](#)]
13. Wang, Y.; Wang, D.; Shi, P.; Omasa, K. Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. *Plant Methods* **2014**, *10*, 36. [[CrossRef](#)] [[PubMed](#)]
14. Ata-Ul-Karim, S.T.; Cao, Q.; Zhu, Y.; Tang, L.; Rehmani, M.I.A.; Cao, W. Non-destructive assessment of plant nitrogen parameters using leaf chlorophyll measurements in rice. *Front. Plant Sci.* **2016**, *7*, 1829. [[CrossRef](#)] [[PubMed](#)]
15. Ata-Ul-Karim, S.T.; Zhu, Y.; Liu, X.; Cao, Q.; Tian, Y.; Cao, W. Comparison of different critical nitrogen dilution curves for nitrogen diagnosis in rice. *Sci. Rep.* **2017**, *7*, 42679. [[CrossRef](#)]
16. Zhang, K.; Jifeng, M.; Yu, W.; Weixing, C.; Yan, Z.; Qiang, C.; Xiaojun, L.; Yongchao, T. Key variable for simulating critical nitrogen dilution curve of wheat: Leaf area ratio-driven approach. *Pedosphere* **2022**, *32*, 463–474. [[CrossRef](#)]
17. Clerget, B.; Bueno, C.; Domingo, A.J.; Layaoen, H.L.; Vial, L. Leaf emergence, tillering, plant growth, and yield in response to plant density in a high-yielding aerobic rice crop. *Field Crops Res.* **2016**, *199*, 52–64. [[CrossRef](#)]
18. Jiang, S.; Du, B.; Wu, Q.; Zhang, H.; Zhu, J. Increasing pit-planting density of rice varieties with different panicle types to improves sink characteristics and rice yield under alternate wetting and drying irrigation. *Food Energy Secur.* **2021**, *12*, e335. [[CrossRef](#)]
19. Zhou, C.; Huang, Y.; Jia, B.; Wang, S.; Dou, F.; Samonte, S.O.P.; Chen, K.; Wang, Y. Optimization of nitrogen rate and planting density for improving the grain yield of different rice genotypes in northeast China. *Agronomy* **2019**, *9*, 555. [[CrossRef](#)]
20. Liu, Q.; Wu, X.; Ma, J.; Chen, B.; Xin, C. Effects of delaying transplanting on agronomic traits and grain yield of rice under mechanical transplantation pattern. *PLoS ONE* **2015**, *10*, e0123330. [[CrossRef](#)]
21. Chen, J.; Zhu, X.; Xie, J.; Deng, G.; Tu, T.; Guan, X.; Yang, Z.; Huang, S.; Chen, X.; Qiu, C. Reducing nitrogen application with dense planting increases nitrogen use efficiency by maintaining root growth in a double-rice cropping system. *Crop J.* **2021**, *9*, 805–815. [[CrossRef](#)]
22. Lecerf, R.; Ceglar, A.; López-Lozano, R.; Van Der Velde, M.; Baruth, B. Assessing the information in crop model and meteorological indicators to forecast crop yield over Europe. *Agric. Syst.* **2019**, *168*, 191–202. [[CrossRef](#)]
23. Gaso, D.V.; Berger, A.G.; Ciganda, V.S. Predicting wheat grain yield and spatial variability at field scale using a simple regression or a crop model in conjunction with Landsat images. *Comput. Electron. Agric.* **2019**, *159*, 75–83. [[CrossRef](#)]
24. Millan, R.; Mougnot, J.; Rabatel, A.; Jeong, S.; Cusicanqui, D.; Derkacheva, A.; Chekki, M. Mapping surface flow velocity of glaciers at regional scale using a multiple sensors approach. *Remote Sens.* **2019**, *11*, 2498. [[CrossRef](#)]
25. Ransom, C.J.; Kitchen, N.R.; Camberato, J.J.; Carter, P.R.; Ferguson, R.B.; Fernández, F.G.; Franzen, D.W.; Laboski, C.A.; Myers, D.B.; Nafziger, E.D. Statistical and machine learning methods evaluated for incorporating soil and weather into corn nitrogen recommendations. *Comput. Electron. Agric.* **2019**, *164*, 104872. [[CrossRef](#)]
26. Kundu, P.P.; Anatharaman, L.; Truong-Huu, T. An empirical evaluation of automated machine learning techniques for malware detection. In Proceedings of the 2021 ACM Workshop on Security and Privacy Analytics, Virtual Event, 28 April 2021; pp. 75–81.
27. Shi, P.; Wang, Y.; Xu, J.; Zhao, Y.; Yang, B.; Yuan, Z.; Sun, Q. Rice nitrogen nutrition estimation with RGB images and machine learning methods. *Comput. Electron. Agric.* **2021**, *180*, 105860. [[CrossRef](#)]

28. Wen, G.; Ma, B.-L.; Vanasse, A.; Caldwell, C.D.; Earl, H.J.; Smith, D.L. Machine learning-based canola yield prediction for site-specific nitrogen recommendations. *Nutr. Cycl. Agroecosyst.* **2021**, *121*, 241–256. [[CrossRef](#)]
29. Li, D.; Miao, Y.; Ransom, C.J.; Bean, G.M.; Kitchen, N.R.; Fernández, F.G.; Sawyer, J.E.; Camberato, J.J.; Carter, P.R.; Ferguson, R.B. Corn nitrogen nutrition index prediction improved by integrating genetic, environmental, and management factors with active canopy sensing using machine learning. *Remote Sens.* **2022**, *14*, 394. [[CrossRef](#)]
30. Chlingaryan, A.; Sukkarieh, S.; Whelan, B. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Comput. Electron. Agric.* **2018**, *151*, 61–69. [[CrossRef](#)]
31. Cao, J.; Zhang, Z.; Tao, F.; Zhang, L.; Luo, Y.; Zhang, J.; Han, J.; Xie, J. Integrating multi-source data for rice yield prediction across china using machine learning and deep learning approaches. *Agric. For. Meteorol.* **2021**, *297*, 108275. [[CrossRef](#)]
32. Berger, K.; Verrelst, J.; Féret, J.-B.; Hank, T.; Woche, M.; Mauser, W.; Camps-Valls, G. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102174. [[CrossRef](#)]
33. Muharam, F.M.; Nurulhuda, K.; Zulkafli, Z.; Tarmizi, M.A.; Abdullah, A.N.H.; Che Hashim, M.F.; Mohd Zad, S.N.; Radhwane, D.; Ismail, M.R. UAV-and Random-Forest-AdaBoost (RFA)-based estimation of rice plant traits. *Agronomy* **2021**, *11*, 915. [[CrossRef](#)]
34. Bean, G.; Kitchen, N.; Camberato, J.; Ferguson, R.; Fernandez, F.; Franzen, D.; Laboski, C.; Nafziger, E.; Sawyer, J.; Scharf, P. Improving an active-optical reflectance sensor algorithm using soil and weather information. *Agron. J.* **2018**, *110*, 2541–2551. [[CrossRef](#)]
35. Aranguren, M.; Castellón, A.; Aizpurua, A. Crop sensor based non-destructive estimation of nitrogen nutritional status, yield, and grain protein content in wheat. *Agriculture* **2020**, *10*, 148. [[CrossRef](#)]
36. Corti, M.; Cavalli, D.; Cabassi, G.; Gallina, P.M.; Bechini, L. Does remote and proximal optical sensing successfully estimate maize variables? A review. *Eur. J. Agron.* **2018**, *99*, 37–50. [[CrossRef](#)]
37. Cui, Z.; Zhang, H.; Chen, X.; Zhang, C.; Ma, W.; Huang, C.; Zhang, W.; Mi, G.; Miao, Y.; Li, X. Pursuing sustainable productivity with millions of smallholder farmers. *Nature* **2018**, *555*, 363–366. [[CrossRef](#)] [[PubMed](#)]
38. Peng, J.; Manevski, K.; Kørup, K.; Larsen, R.; Andersen, M.N. Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. *Field Crops Res.* **2021**, *268*, 108158. [[CrossRef](#)]
39. Li, J.; Xu, M.; Xin, J.; Duan, J.; Ren, Y.; Li, D.; Huang, J.; Shen, H.; Zhang, H. Spatial and Temporal Characteristics of Basic Soil Productivity in China. *Sci. Agric. Sin.* **2016**, *49*, 1510–1519.
40. China Rice Data Center. National Rice Data Center Variety Profile. Available online: <https://www.ricedata.cn/variety/index.htm> (accessed on 1 January 2022).
41. Nelson, D.W.; Sommers, L. Determination of total nitrogen in plant material 1. *Agron. J.* **1973**, *65*, 109–112. [[CrossRef](#)]
42. Rodriguez, J.D.; Perez, A.; Lozano, J.A. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *32*, 569–575. [[CrossRef](#)]
43. Jiang, R.; He, W.; Zhou, W.; Hou, Y.; Yang, J.; He, P. Exploring management strategies to improve maize yield and nitrogen use efficiency in northeast China using the DNDC and DSSAT models. *Comput. Electron. Agric.* **2019**, *166*, 104988. [[CrossRef](#)]
44. Bo, Y.; He, H.-B.; Xu, H.-C.; Zhu, T.-Z.; Tao, L.; Jian, K.; You, C.-C.; Zhu, D.; Wu, L.-Q. Determining nitrogen status and quantifying nitrogen fertilizer requirement using a critical nitrogen dilution curve for hybrid indica rice under mechanical pot-seedling transplanting pattern. *J. Integr. Agric.* **2021**, *20*, 1474–1486.
45. Yao, X.; Ata-Ul-Karim, S.T.; Zhu, Y.; Tian, Y.; Liu, X.; Cao, W. Development of critical nitrogen dilution curve in rice based on leaf dry matter. *Eur. J. Agron.* **2014**, *55*, 20–28. [[CrossRef](#)]
46. Hu, Y.-J.; Pei, W.; Zhang, H.-C.; Dai, Q.-G.; Huo, Z.-Y.; Ke, X.; Hui, G.; Wei, H.-Y.; Guo, B.-W.; Cui, P.-Y. Comparison of agronomic performance between inter-sub-specific hybrid and inbred japonica rice under different mechanical transplanting methods. *J. Integr. Agric.* **2018**, *17*, 806–816. [[CrossRef](#)]
47. Landau, S.; Mitchell, R.; Barnett, V.; Colls, J.; Craigon, J.; Payne, R. Response to “Comments on” Testing winter wheat simulation models predictions against observed UK grain yields. *Agric. For. Meteorol.* **1999**, *96*, 163–164. [[CrossRef](#)]
48. Hernandez, J.; Lobos, G.A.; Matus, I.; Del Pozo, A.; Silva, P.; Galleguillos, M. Using ridge regression models to estimate grain yield from field spectral data in bread wheat (*Triticum aestivum* L.) grown under three water regimes. *Remote Sens.* **2015**, *7*, 2109–2126. [[CrossRef](#)]
49. Shahhosseini, M.; Martinez-Feria, R.A.; Hu, G.; Archontoulis, S.V. Maize yield and nitrate loss prediction with machine learning algorithms. *Environ. Res. Lett.* **2019**, *14*, 124026. [[CrossRef](#)]
50. Lemaire, G.; Ciampitti, I. Crop mass and N status as prerequisite covariables for unraveling nitrogen use efficiency across genotype-by-environment-by-management scenarios: A review. *Plants* **2020**, *9*, 1309. [[CrossRef](#)] [[PubMed](#)]
51. Zhao, B.; Zhang, Y.; Duan, A.; Liu, Z.; Xiao, J.; Liu, Z.; Qin, A.; Ning, D.; Li, S.; Ata-Ul-Karim, S.T. Estimating the growth indices and nitrogen status based on color digital image analysis during early growth period of winter wheat. *Front. Plant Sci.* **2021**, *12*, 619522. [[CrossRef](#)] [[PubMed](#)]
52. Liu, X.-j.; Qiang, C.; Yuan, Z.-f.; Xia, L.; Wang, X.-l.; Tian, Y.-c.; Cao, W.-x.; Yan, Z. Leaf area index based nitrogen diagnosis in irrigated lowland rice. *J. Integr. Agric.* **2018**, *17*, 111–121. [[CrossRef](#)]
53. Xu, H.; He, H.; Yang, K.; Ren, H.; Zhu, T.; Ke, J.; You, C.; Guo, S.; Wu, L. Application of the Nitrogen Nutrition Index to Estimate the Yield of Indica Hybrid Rice Grown from Machine-Transplanted Bowl Seedlings. *Agronomy* **2022**, *12*, 742. [[CrossRef](#)]

54. Forkuor, G.; Hounkpatin, O.K.; Welp, G.; Thiel, M. High resolution mapping of soil properties using remote sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. *PLoS ONE* **2017**, *12*, e0170478. [[CrossRef](#)] [[PubMed](#)]
55. Ata-Ul-Karim, S.T.; Liu, X.; Lu, Z.; Zheng, H.; Cao, W.; Zhu, Y. Estimation of nitrogen fertilizer requirement for rice crop using critical nitrogen dilution curve. *Field Crops Res.* **2017**, *201*, 32–40. [[CrossRef](#)]
56. Patel, M.K.; Padarian, J.; Western, A.W.; Fitzgerald, G.J.; McBratney, A.B.; Perry, E.M.; Suter, H.; Ryu, D. Retrieving canopy nitrogen concentration and aboveground biomass with deep learning for ryegrass and barley: Comparing models and determining waveband contribution. *Field Crops Res.* **2023**, *294*, 108859. [[CrossRef](#)]
57. Huang, M.; Chen, J.; Cao, F.; Zou, Y. Increased hill density can compensate for yield loss from reduced nitrogen input in machine-transplanted double-cropped rice. *Field Crops Res.* **2018**, *221*, 333–338. [[CrossRef](#)]
58. Zhou, C.; Huang, Y.; Jia, B.; Wang, Y.; Wang, Y.; Xu, Q.; Li, R.; Wang, S.; Dou, F.J.A. Effects of cultivar, nitrogen rate, and planting density on rice-grain quality. *Agronomy* **2018**, *8*, 246. [[CrossRef](#)]
59. Bai, Y.; Gao, J. Optimization of the nitrogen fertilizer schedule of maize under drip irrigation in Jilin, China, based on DSSAT and GA. *Agric. Water Manag.* **2021**, *244*, 106555. [[CrossRef](#)]
60. Puntel, L.A.; Sawyer, J.E.; Barker, D.W.; Dietzel, R.; Poffenbarger, H.; Castellano, M.J.; Moore, K.J.; Thorburn, P.; Archontoulis, S.V. Modeling long-term corn yield response to nitrogen rate and crop rotation. *Front. Plant Sci.* **2016**, *7*, 1630. [[CrossRef](#)] [[PubMed](#)]
61. Alotaibi, K.D.; Cambouris, A.N.; St. Luce, M.; Ziadi, N.; Tremblay, N. Economic optimum nitrogen fertilizer rate and residual soil nitrate as influenced by soil texture in corn production. *Agron. J.* **2018**, *110*, 2233–2242. [[CrossRef](#)]
62. Luo, Z.; Liu, H.; Li, W.; Zhao, Q.; Dai, J.; Tian, L.; Dong, H. Effects of reduced nitrogen rate on cotton yield and nitrogen use efficiency as mediated by application mode or plant density. *Field Crops Res.* **2018**, *218*, 150–157. [[CrossRef](#)]
63. Zhang, Y.; Wang, H.; Lei, Q.; Luo, J.; Lindsey, S.; Zhang, J.; Zhai, L.; Wu, S.; Zhang, J.; Liu, X. Optimizing the nitrogen application rate for maize and wheat based on yield and environment on the Northern China Plain. *Sci. Total Environ.* **2018**, *618*, 1173–1183. [[CrossRef](#)]
64. Yan, F.; Zhang, F.; Fan, X.; Fan, J.; Wang, Y.; Zou, H.; Wang, H.; Li, G. Determining irrigation amount and fertilization rate to simultaneously optimize grain yield, grain nitrogen accumulation and economic benefit of drip-fertigated spring maize in northwest China. *Agric. Water Manag.* **2021**, *243*, 106440. [[CrossRef](#)]
65. Wang, M.; Wang, H.; Hou, L.; Zhu, Y.; Zhang, Q.; Chen, L.; Mao, P. Development of a critical nitrogen dilution curve of Siberian wildrye for seed production. *Field Crops Res.* **2018**, *219*, 250–255. [[CrossRef](#)]
66. Huang, G.; Zhang, Y.; Zhang, S.; Zhang, J.; Hu, F.; Li, F. Density-Dependent Fertilization of Nitrogen for Optimal Yield of Perennial Rice. *Agronomy* **2022**, *12*, 1698. [[CrossRef](#)]
67. Li, F.; Li, D.; Elsayed, S.; Hu, Y.; Schmidhalter, U. Using optimized three-band spectral indices to assess canopy N uptake in corn and wheat. *Eur. J. Agron.* **2021**, *127*, 126286. [[CrossRef](#)]
68. Elsayed, S.; El-Hendawy, S.; Dewir, Y.H.; Schmidhalter, U.; Ibrahim, H.H.; Ibrahim, M.M.; Elsherbiny, O.; Farouk, M. Estimating the leaf water status and grain yield of wheat under different irrigation regimes using optimized two-and three-band hyperspectral indices and multivariate regression models. *Water* **2021**, *13*, 2666. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.