



Article **Prediction of Pea (***Pisum sativum* **L.) Seeds Yield Using Artificial Neural Networks**

Patryk Hara ¹, Magdalena Piekutowska ² and Gniewko Niedbała ^{3,*}

- ² Department of Geoecology and Geoinformation, Institute of Biology and Earth Sciences, Pomeranian University in Słupsk, 27 Partyzantów St., 76-200 Słupsk, Poland
- ³ Department of Biosystems Engineering, Faculty of Environmental and Mechanical Engineering,
- Poznań University of Life Sciences, Wojska Polskiego 50, 60-627 Poznań, Poland

* Correspondence: gniewko.niedbala@up.poznan.pl

Abstract: A sufficiently early and accurate prediction can help to steer crop yields more consciously, resulting in food security, especially with an expanding world population. Additionally, prediction related to the possibility of reducing agricultural chemistry is very important in an era of climate change. This study analyzes the performance of pea (*Pisum sativum* L.) seed yield prediction by a linear (MLR) and non-linear (ANN) model. The study used meteorological, agronomic and phytophysical data from 2016–2020. The neural model (N2) generated highly accurate predictions of pea seed yield—the correlation coefficient was 0.936, and the RMS and MAPE errors were 0.443 and 7.976, respectively. The model significantly outperformed the multiple linear regression model (RS2), which had an RMS error of 6.401 and an MAPE error of 148.585. The sensitivity analysis carried out for the neural network showed that the characteristics with the greatest influence on the yield of pea seeds were the date of onset of maturity, the date of harvest, the total amount of rainfall and the mean air temperature.

Keywords: pea; seeds yield prediction; ANN; MLR; sensitivity analysis



Citation: Hara, P.; Piekutowska, M.; Niedbała, G. Prediction of Pea (*Pisum sativum* L.) Seeds Yield Using Artificial Neural Networks. *Agriculture* 2023, *13*, 661. https:// doi.org/10.3390/agriculture13030661

Academic Editors: Raul Morais dos Santos and Bing Liu

Received: 14 February 2023 Revised: 26 February 2023 Accepted: 10 March 2023 Published: 12 March 2023



Copyright: © 2023 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

There are many challenges facing modern agriculture. The priority is to increase food production while minimizing environmental impact [1,2]. This task seems particularly difficult in the face of climate change and the occurrence of increasingly frequent extreme weather events, which pose a serious threat to crop yields [3]. It is assumed that about 67% of crop variability is governed by the weather conditions that prevail throughout the crop growing season, with 33% governed by other factors, such as agrotechnology or habitat conditions [4]. Therefore, early and accurate forecasting of crop yields is becoming increasingly important [5]. Being able to estimate yields a few weeks before harvest allows an appropriate strategy to be taken for the pricing of agricultural products. Yield prediction can also be a useful tool for decision makers in regulating both exports in case of surpluses and imports in times of agricultural commodity shortages [6]. In addition, early information on yields can help farmers in terms of work planning and storage space selection [7]. This knowledge can also help to improve the profitability of agricultural production by optimizing the number of crop protection and/or fertilization treatments. Lower usage of these products leads to a reduction in total labor inputs and on-farm energy inputs [8,9]. In the final balance, these factors contribute to increased labor productivity, conservation of natural resources and increased farm profitability through lower production costs [10]. Accurate and early prediction of crop yields also plays an important role in global food security by providing valuable information to various stakeholders (farm owners, agronomists, etc.) [11,12].

The need for accurate and timely predictive models for agricultural crop yield has led to a growing interest in this topic from the scientific community [13–18]. However,

¹ Agrotechnology, Jagiellonów 4, 73-150 Łobez, Poland

developing a predictive model is not an easy task [19]. The issue is extremely complex due to the multitude of factors affecting crop yield. The most commonly cited determinants include genotype and weather conditions, including rainfall, sunshine, minimum and maximum air temperature, habitat conditions (soil pH, soil nutrient abundance, etc.), agronomics and the interactions of these factors [20–23]. Many different approaches are used in yield forecasting, and each method has strengths as well as limitations [24]. One such method is multiple linear regression (MLR). MLR, as a statistical tool, is able to predict yield based on, among other things, agronomic data. However, the effectiveness of this method is often questioned due to its low prediction accuracy [25]. The major disadvantage of MLR models is that they are not appropriate for explaining non-linear and complex relationships between yield and the factors that influence yield [26].

With the development of information technology, modern mathematical algorithm techniques such as machine learning (ML) have begun to be applied. The possibility of using models based on artificial intelligence has contributed to an increase in accurate forecasting of random and non-linear issues [27]. This feature has made machine learning the method most commonly used in yield modeling [28,29]. In addition to its high prediction quality, ML is able to identify patterns in datasets and reveal complex relationships between independent variables [30]. Additionally, the advantage of ML over traditional linear regression methods lies in its ability to use, as explanatory variables, two or more spectral variables from satellite imagery [31]. The inclusion of these data in yield modeling is becoming an increasingly common practice due to further improvements in prediction quality and the prospect of capturing new correlations between these factors and crop yield [32]. Furthermore, in machine learning-based models, it is possible to use linguistic variables without having to code them in advance, as is the case with regression models [9,33]. The accuracy of yield estimation that is achieved by machine learning methods means that these models require large datasets from a variety of sources. In the case of a small number of predictors, the proper calculation of yield variability by ML usually suffers from a large prediction error [34]. Other limitations in the use of ML are that some methods require computationally powerful equipment and that analysis time is much longer than it is for multivariate linear regression [35,36]. Machine learning models are also sensitive to significant correlations between independent characteristics. For this reason, the dataset that is fed into the model often requires prior preparation, and additional statistical analyses may need to be performed to capture these correlations [37]. Some of the most successful machine learning techniques are support vector machines (SVMs), convolutional neural networks (CNNs), random forest (RF), k-nearest neighbors (kNNs) and artificial neural networks (ANNs) [38-41].

ANNs are a mathematical tool that can create a non-linear representation of the connections between the explained variable and the input variables [42]. ANNs are, to some extent, inspired by the functioning of parts of the real (biological) nervous system [43]. However, the connection patterns of neurons in artificial neural networks are chosen arbitrarily and are not a model of actual neural structures. ANNs as a computer tool are distinguished by their ability to solve practical problems in a computerized manner without prior mathematical formalization [44]. Another advantage is that it is not necessary to refer to any theoretical assumptions about the problem being solved when working with neural networks. Even the assumption of causal relationships between exploratory and explanatory features need not be enforced [45]. The computations performed by the ANN are performed in parallel. The artificial neurons that make up the network perform their computational tasks simultaneously. This makes the network capable of solving the problem under analysis in a short period of time. However, the more complex the problem the neural network investigates, the more time it takes to find the right solution [46]. The most characteristic feature of artificial neural networks is the ability to learn from examples and the ability to self-generalize the acquired knowledge (generalization) [47]. The threat to generalization is overlearning. An overlearned network excessively adapts the acquired knowledge to irrelevant details of specific learning cases [48].

One commonly used ANN model is the multilayer perceptron (MLP) [17,18,49–51]. It is a fully connected unidirectional neural network [7] that typically consists of three layers: an input layer, at least one hidden layer consisting of sigmoidal neurons and an output layer consisting of sigmoidal or linear neurons. The back-propagation method is the most commonly used technique for learning MLP networks [52]. This method is based on the concept of correcting, at each stage of learning, the values of the weights based on the evaluation of the error made by each neuron during the learning of the network [53].

The present work is a continuation of the authors' previous research [54], which aimed to determine the effectiveness of linear (MLR) and non-linear (MLP) models in predicting the protein content of *Pisum sativum* L. pea seeds. The current study focused on the possibility of predicting the seed yield of general pea seeds using ANNs, and MLR was used as a comparative model. In addition, the study aimed to test three hypotheses: (i) the artificial neural networks model is an effective tool in predicting pea seed yield 20 days before harvest; (ii) five-year field trials under a variety of experimental conditions allow for the construction of a model predicting pea yield; and (iii) neural networks can predict yield more accurately than the MLR model.

2. Materials and Methods

This research was carried out between 2016 and 2020 at the Stations and Experimental Plants of the Research Center for Cultivar Testing (COBORU). The mission of COBORU is to stimulate innovation in plant breeding and seed science and to support the implementation of diverse progress into agricultural practice [55]. The work of this unit is focused, among other things, on research into the distinctiveness, uniformity and stability of crop varieties in Poland. In addition, COBORU is involved in conducting field research on the assessment of the cultivation and use value of agricultural crops. These studies are conducted under conditions as close as possible to production conditions. The results of the conducted experiments make it possible to determine whether a given variety can be entered into the National List of Varieties [56].

The experimental plots were located in Poland at the following locations: Bezek (N 51°12'6.722" E 23°16'7.656"), Głębokie (N 52°38'33.18" E 18°26'16.26"), Kaweczyn (N 52°10'15.157" E 20°20'49.328"), Krzyżewo (N 53°1'33. 535" E 22°45'28.438"), Pawłowice (N 50°27'14.049" E 18°29'28.912"), Radostowo (N 53°59'20.566" E 18°44'41.429") and Sulejów (N 51°21′8.03″ E 19°52′7.517″). The experiments were situated in locations that are optimal for pea cultivation in terms of habitat. These localities are characterized by a temperate warm climate, with average monthly air temperatures ranging from -5.0 to -2.0 °C in January and 16 to 18 °C in July. Average annual precipitation is in the range of 550–800 mm [57]. According to Polish soil classification, clay soils of classes II-IIIb prevail in these localities. The data for the construction of the models are official data, coming from a variety of COBORU tests, and are acceptable to all authorities related to agricultural production in Poland. The nature of the experiments and the way in which they were carried out are recorded in the methodology [58], which is a set of experimental concepts and guidelines. This ensures that all research assumptions are met. The research was conducted in the same way at each COBORU point. The meteorological data were obtained from the archive database of the Institute of Meteorology and Water Management at the National Research Institute. A detailed description of the conduct of the experiments, the acquisition of the dataset and the sources of these data were previously described by the authors of this paper [54]. The construction of an ANN (N2) and MLR (RS2) model was performed based on 11 general purpose pea cultivars: Arwena, Astronaute, Batuta, Mecenas, Medyk, Mentor, Olimp, Spot, Starski, Tarchalska and Tytus.

2.1. Construction of the Database

The first and most important step in the construction of linear and non-linear models is the appropriate selection of input variables. The importance of this step is due to the fact that the chosen input parameters directly affect the performance of the resulting models [37].

The input variables shown in Table 1 were used to build the N2 and RS2 model. The output variable was pea seed yield expressed in $t \cdot ha^{-1}$. The dataset consisted of 1155 cases/plots. Each plot was a separate case for model building. All cases that formed the dataset were divided into two sets: A and B. Data from 1040 plots were assigned to set A, while set B was created from the remaining 115 cases and was used for model validation.

Table 1. The structure and scope of the independent variables used in the construction and verification of the N2 and RS2 models.

Symbol	Unit of Measure	Description of the Variable	The Scope of Data		
Independent Variables					
RAIN	mm	Rainfall in the period from sowing to 14 July	96.9–312.4		
SUN	h	Sum of insolation that occurred in the period from sowing to 14 July	630.5–1051.5		
TEMP	°C	Average daily air temperature in the period from sowing to 14 July	11.0–17.5		
N_F	kg/ha	Amount of nitrogen introduced into the soil with mineral fertilizers	10–90		
P2O5_F	kg/ha	Amount of phosphorus incorporated into the soil with mineral fertilizers	0–80		
K2O_F	kg/ha	Amount of potassium introduced into the soil with mineral fertilizers	0–119		
SOWI	days	Date of sowing of field peas—defined as number of days since the beginning of the year	83–102		
P_EMER	days	Pea crop emergence—defined as number of days since the beginning of the year	96–133		
HAR	days	Date of harvesting of field pea plants—defined as the number of days from 1 January	184–221		
FLOWE	days	Flowering onset date—number of days from the beginning of the year	126–169		
INI_MA	days	Maturity onset date—defined as the number of days from 1 January	167–211		
TECH_M	days	Technical maturity date—number of days since the beginning of the year	171–216		
P_HIG	cm	Height of plants	43-156		
WEGW	days	Number of plant growing days	87–137		
PH	-	Soil reaction (pH)	5.5–7.5		
P2O5_C	Range from 0 to 4 *	Phosphorus (V) oxide content of the soil	0–4		
K2O_C	Range from 0 to 4 *	Potassium oxide content of the soil	0–4		
MGO_C	Range from 0 to 4 *	Magnesium oxide content of the soil	0–4		
GEN	Feature coded 101 to 111	Variety of peas	-		
Dependent Variable					
YIELD	t∙ha ^{−1}	Pea seed yield	2.30-8.02		

* The range from 0 to 4 refers to the nutrient abundance of the soil. A value of 0 indicates very low abundance, 1 indicates low abundance, 2 indicates medium abundance, 3 indicates high abundance and 4 indicates very high abundance. Range from 0 to 4.

2.2. Construction of the N2 Model

In the present study, it was assumed that the forecast of pea seed yield would be made before harvesting [59], i.e., 14 July. The forecast date was selected based on the dominance of the onset of maturity of the pea varieties included in the dataset. The analysis of the dataset showed that the harvesting of peas of general use varieties was most often performed on 3 August. Therefore, the obtained linear and non-linear model predicted the yield 20 days before the harvest of peas grown under experimental conditions.

The construction of N2 models consisted of input variables being repeatedly provided to the network [60]. A total of 10,000 neural networks were tested using an automatic network designer. Different ANN model structures were analyzed, including variations in the number of neurons in the hidden layer. This method selected a model with an MLP architecture of 19:19-24-1:1 (Figure 1). The multilayer perceptron is a type of ANN widely recommended in works on similar topics due to its high potential for non-linear function estimation [18,61–63]. The main advantage of MLPs is the ability to discriminate data that cannot be linearly separated [7].



Figure 1. Network structure for the N2 model.

The optimization of the artificial neural network structure was performed by obtaining the minimum validation error. The selection of the neural network model was also guided by the size of the training and test set errors and other important quality parameters, as shown in Table 2. In this research, set A was divided into three subsets: learning, test and validation. This division is common in the development of predictive models using ANNs [5,59,64]. In the present study, 50% of the records (or 520 cases) were assigned to the learning subset. The test and validation subsets consisted of the same number of objects, i.e., 260, each representing 25% of the cases from the entire A set. The construction of the N2 model was carried out using Statistica v7.1 (TIBCO Software Inc., Palo Alto, CA, USA).

Subsets	Teaching	Validation	Testing
Size of error	0.0556	0.0590	0.0679
Quality	0.3576	0.3645	0.4311
	Epochs of	of learning	
Back-propagation method of error		10	00
Coupled gradients method		110	lb *
b (best)—the best result in t	he indicated learning epoc	h.	

Table 2. Quality parameters and error rates of subsets and the number of learning epochs of neural networks.

o (ocor) the ocor result in the indicated rearring ep

2.3. Construction of the RS2 Model

Due to their simplicity, multivariate linear regression models are commonly used in the prediction of agricultural crop yields [65]. MLR models the combination of a dependent trait and two or more independent traits by creating a linear equation to the observed data [66]. The value of the explanatory variable (Y) is related to the value of the explanatory variables (X) according to Equation (1) [62]:

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + \dots + b_p \cdot X_p + \varepsilon,$$
 (1)

where Y is the dependent variable (explained variable), $X_1, X_2 \dots X_p$ represents the independent variables (explanatory variable), $b_0, b_1, b_2 \dots b_p$ represents equation parameters and ε denotes the random component (rest of the model).

For the purpose of this work, an MLR model (stepwise progressive) was built based on the explanatory variables presented in Table 1. The procedure for building the RS2 model was similar to that for the N2 model. The computational analysis took eighteen steps. All the steps involved in building and verifying the RS2 model were performed, as with the N2 models, in Statistica v7.1 (TIBCO Software Inc., Palo Alto, CA, USA).

2.4. Evaluation Criteria for the N2 and RS2 Models

Six performance criteria (global relative error of model approximation (RAE), root mean square error (RMS), mean absolute error (MAE), mean absolute percentage error (MAPE), maximum error determined for the whole model (MAX) and maximum percentage error (MAXP)) were used to evaluate the resulting predictive models [54]. In order to calculate the values of these errors, a set B was required, which was used to determine the difference between the predicted and observed values. The magnitudes of these errors were calculated from the equations below:

RAE =
$$\sqrt{\frac{\sum_{i=1}^{n} (y_i - y'_i)^2}{\sum_{i=1}^{n} (y_i)^2}}$$
, (2)

RMS =
$$\sqrt{\frac{\sum_{i=1}^{n} (y_i - y'_i)^2}{n}}$$
 (3)

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y'_i|$$
(4)

MAPE =
$$\frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y'_i}{y_i} \right| \cdot 100\%$$
 (5)

$$MAX = \max_{i} \cdot |y_{i} - y'_{i}|$$
(6)

$$MAXP = \max_{i} \left| \frac{y_{i} - y'_{i}}{y_{i}} \right| \cdot 100\%$$
(7)

where *n* is the number of observations, y_i is the actual values and y'_i is the predictive values obtained with the model.

2.5. Sensitivity Analysis of the Neural Network

The final stage in the construction of predictive models based on ANNs is sensitivity analysis of the neural network. This stage consists of differentiating the independent variables in terms of their influence on the dependent variable. The method of calculating and interpreting the results obtained from the sensitivity analysis has been discussed in previous works by the authors of this paper [17,18,54].

3. Results

3.1. Overall Assessment of the Predictive Quality of the N2 and RS2 Models

Building predictive models based on artificial neural networks requires partitioning of the dataset. In this paper, the dataset was divided into three subsets: learning, validation and testing. The learning set contains both input and output data as patterns of valid signals. Based on these, the learning algorithm confronts the actual behavior of the network. The validation set is used indirectly in the learning process. Its task is to participate in periodic validation during the learning of the model, which prevents the occurrence of network overfitting [67]. The test set, on the other hand, is intended for one-time control after the training is completed. In general, this procedure is aimed at checking whether there was any loss of network generalization ability during training that may have resulted from coincidence, despite cyclical internal validation. Table 2 shows the error sizes and the quality of each subset. From this, we can observe that the learning set had the smallest error size (0.0556). The test set, on the other hand, was characterized by the largest error among the analyzed subsets. The error value for this subset was 0.0679. A different relationship can be observed in the case of subset quality. The test set was characterized by the highest value of this feature (0.4311), and the learning set by the lowest (0.3576). The validation set took an intermediate position (0.3645).

The N2 model was learned using two methods: back-propagation of the error and the coupled gradients method. The point at which there is an increase in error for the validation set is the signal to stop training the neural network and recover the best weight from the epoch that preceded the start of the error increase. In the case at hand, the first learning method lasted for 100 epochs, and by continuing the learning process with the coupled gradients method, it was possible to obtain the best result, which was achieved at 110 epochs.

The obtained N2 model predicting pea seed yield was characterized by a relatively low mean error value, which was about 0.015 (Table 3). The model also obtained a small mean absolute error, which did not exceed the value of 0.305. In turn, the correlation coefficient reached a relatively high value (0.936). In developing predictive models, it is important that the model built is characterized by low error magnitudes and a high correlation coefficient value, as only such a model will be able to accurately predict the dependent variable.

Table 3. Qualitative measures of the N2 model.

Quality Parameter	Value
Average	4.504
Standard deviation	1.106
Average error	0.015
Error deviation	0.389
Average absolute error	0.305
Deviation quotient	0.352
Correlation coefficient r	0.936

Multivariate linear regression analysis showed that the input variables that were not statistically significant ($\alpha = 0.05$) were the date of harvest (HAR), the date of plant technical maturity (TECH_M) and the dose of potassium brought into the soil with mineral fertilizer (K2O_F). Based on the results in Table 4, the form of the MLR equation was determined:

 $\begin{aligned} \text{YIELD} &= 0.215 \times \text{P2O5}_{C} - 0.23 \times \text{N}_{F} - 0.089 \times \text{P}_{E}\text{MER} + 0.123 \text{ INI}_{M}\text{A} - 0.007 \times \text{TEMP} - 0.007 \times \text{RAIN} - \\ 0.003 \times \text{SUN} + 0.022 \times \text{P2O5}_{F} + 0.481 \times \text{PH} + 0.169 \times \text{K2O}_{C} - 0.040 \times \text{FLOWE} - 0.150 \times \text{MGO}_{C} - 0.031 \times \\ \text{GEN} + 0.005 \times \text{P}_{H}\text{HG} + 0.034 \times \text{WEGW} \end{aligned}$ (8)

3.2. Evaluation of Neural Network Sensitivity Analysis

The purpose of sensitivity analysis is to identify the independent variables that most influenced the dependent trait, pea seed yield. Based on the study, it can be observed that the onset of pea plant maturity (INI_MA) influenced the yield to the greatest extent (Table 5). This feature was ranked 1, and not including it in the N2 models would increase the cumulative error by a factor of 2378. The feature that received a rank of 2 was harvest date (HAR). Not including this variable in the model would increase the cumulative error by about 1.677 times. The variables ranked 3 and 4 were rainfall (RAIN) and mean air temperature (TEMP) calculated from sowing to 14 July. The absence of these variables in the N2 models would increase the cumulative error by 1.575 and 1.471 times, respectively.

Figure 2 shows a scatter plot of observed versus predicted values. From it, it can be concluded that the N2 model was characterized by a good level of prediction of pea seed yield, as evidenced by the relatively high value of the coefficient of determination (R^2), which was about 0.84. A much lower value of this indicator was obtained for the RS2 model (Figure 3). The R^2 coefficient did not exceed a value of 0.58, indicating that the response of the model is strongly discrepant with the observed values. The resulting model has virtually no ability to adequately represent the relationships characteristic of the issue under consideration.

Factor	MLR: r = 0.7656 R ² = 0.5788 Standard Error of Estimate = 0.7184					
-	Beta	Standard Error Beta	b	Standard Error b	р	Significance
Free Term	-	_	-2.207	2.018	0.274282	_
HAR	-0.086	0.136	-0.010	0.017	0.526117	_
P2O5_C	0.177	0.026	0.215	0.031	0.000000	+
N_F	-0.1166	0.030	-0.012	0.003	0.000108	+
P_EMER	-0.480	0.046	-0.089	0.009	0.000000	+
INI_MA	1.027	0.124	0.123	0.015	0.000000	+
TEMP	0.398	0.080	0.283	0.057	0.000001	+
RAIN	-0.343	0.030	-0.007	0.001	0.000000	+
SUN	-0.225	0.029	-0.003	0.000	0.000000	+
P2O5_F	0.370	0.040	0.022	0.002	0.000000	+
PH	0.209	0.029	0.481	0.068	0.000000	+
K2O_C	0.143	0.029	0.169	0.035	0.000001	+
FLOWE	-0.199	0.040	-0.040	0.008	0.000001	+
MGO_C	-0.144	0.031	-0.150	0.032	0.000004	+
GEN	-0.089	0.021	-0.031	0.007	0.000017	+
P_HIG	0.077	0.030	0.005	0.002	0.010286	+
TECH_M	-0.200	0.114	-0.024	0.013	0.078386	—
WEGW	0.358	0.167	0.034	0.0158	0.032558	+
K2O_F	0.067	0.038	0.003	0.002	0.081382	_

 Table 4. Results of multiple linear regression (MLR) analysis.

Determination of the level of statistical significance: - non-significant; + significant for $\alpha = 0.05$.

Table 5. Results of the neural network sensitivity analysis.

Variable	Quotient	Rank
INI_MA	2.378	1
HAR	1.677	2
RAIN	1.575	3
TEMP	1.471	4
P_EMER	1.468	5
MGO_C	1.395	6
SOWI	1.387	7
K2O_C	1.356	8
P2O5_F	1.333	9
WEGE	1.261	10
P_HIG	1.170	11
PH	1.136	12
TECH_M	1.129	13
K2O_F	1.112	14
P2O5_C	1.110	15
GEN	1.079	16
SUN	1.052	17
FLOWE	1.052	18
N-F	1.045	19



Figure 2. Correlation diagram of observed values against predicted values for the N2 model.



Figure 3. Correlation diagram of observed values against predicted values for the RS2 model.

The sensitivity analysis of the N2 model shows that some of the most important variables affecting pea yield were the date of onset of plant maturity (INI_MA) and the date of harvest (HAR). The relationship of these variables is shown in Figure 4. The plants reaching the maturity stage results in low seed yield. The same is true for early harvesting, which can also result in low yields. Higher yields were achievable when plants reached the onset of maturity later and when the harvest date was later.



Figure 4. Response surface for yield size and two variables: HAR (harvest date) and INI_MA (start of crop maturity).

The relationship between pea yield and seed harvest date (HAR) and mean daily air temperature (TEMP) is shown in Figure 5, from which we can observe that the yield increased with increasing mean daily air temperature. A later harvest date also contributed to higher yields. At low TEMP, harvesting too early resulted in low crop efficiency. In addition, it can be observed from Figure 5 that temperature was the characteristic that most determined harvest date. An average daily air temperature of 18 °C allowed the seeds to be harvested around 29 June (180 days from the beginning of the year).



Figure 5. Response surface for yield size and two variables: HAR (harvest date) and TEMP (mean daily air temperature).

Based on the sensitivity analysis of the neural network, it was possible to determine the relationship of the independent variables of onset of maturity (INI_MA) and mean daily temperature (TEMP) in relation to seed yield (Figure 6). Plants yielded best when they reached the onset of maturity stage later (215 days) and when the average air temperature exceeded 17 °C. When peas reached the onset of maturity stage at 165 days (counted from the beginning of the year), plants were characterized by low yields (about 2 t \cdot ha⁻¹). An increase in the TEMP trait contributed to an increase in plant yield. However, when the onset of maturity was reached early, this increase was insignificant and pea seed yield did not exceed 3 t \cdot ha⁻¹.



Figure 6. Response surface for yield size and two variables: INI_MA (start of crop maturity) and TEMP (mean daily air temperature).

3.3. Comparative Analysis of N2 and RS2 Models Based on Model Evaluation Criteria

The constructed models were verified for their validity. For this purpose, basic quality criteria were used, the values of which are shown in Table 6. The N2 model achieved a root mean square error (RAE) of 0.094 and the maximum percentage error (MAPE) was around 7.98. Much larger error values were obtained for the RS2 model. The RMS error was determined to be 6.401 and the MAPE reached a value of 148.585. Such high errors make the multiple linear regression method an unsuitable tool for forecasting pea seed yield. The strongly non-linear relationships affecting pea yield cause the linear method to mispredict the dependent variable.

Error Type	N2 Model	RS2 Model
RAE	0.094	1.361
RMS	0.443	6.401
MAE	0.347	6.361
MAPE	7.976	148.585
MAX	1.398	7.739
MAXP	48.050	237.384

Table 6. The quality of the generated neural models.

4. Discussion

Yield prediction methods have been used extensively in a number of works and have built models predicting the yields of maize, potato, winter wheat and orchard fruit, among others [68–73]. Specialized equipment, such as drones equipped with multispectral cameras, has been used to build some models in order to obtain information on crop characteristics. However, these devices can often be very expensive, and it is additionally necessary to have adequate knowledge of their operation. Consequently, models based on field imaging can be very difficult for agricultural producers and their application in agricultural practice may be limited. The proposed N2 and RS2 models were built using weather, agronomic and phytophenological data. Analysis of the prediction quality of the models showed that multiple linear regression was ineffective in estimating pea seed yield under the experimental conditions. The obtained MAPE error of 148.585 significantly exceeded the prediction accuracy threshold, disqualifying the obtained model as a suitable tool in yield prediction. According to Peng et al. [15], a predictive model that achieves an MAPE error greater than 30% is characterized by a poor representation of the predicted and observed values of the compound and should therefore be discarded. When MAPE is <10%, the model exhibits an excellent degree of fit. Such a low mean absolute percentage error was obtained for the N2 model based on ANNs. This model had an MAPE error of 7.976 (Table 6). It should be noted, however, that there is no appropriate comparative method for ANNs, and such models are therefore mainly compared with classical regression analyses worldwide. Such comparisons have been made, among others, by Kumari et al. [74], who predicted the yield of Indian nickel (Cajanus cajan—a bean crop) using a two-layer feed-forward neural network and an MLR model. The study was conducted in the Varanasi region (India) and the input data used were for the 1985–1986 and 2011–2012 periods. Five weather characteristics were used to build the models, i.e., minimum and maximum temperature, rainfall, and maximum and minimum relative humidity. The study conducted showed that the ANN model outperformed the MLR model in the prediction of Indian nickel yield. The RMS error for the ANN model was 299.93 kg·ha⁻¹, while the MLR model had an error magnitude of 884.02 kg \cdot ha⁻¹. Artificial neural networks and multiple regression models were also used for the prediction of crescent beans (Phaseolus lunatus L.) [75]. The study was conducted in the northeastern part of Brazil, and independent variables used included the date of flowering onset, the date of pod maturity onset and pod length. The analysis showed that the MLP model forecasted yield more accurately compared to the MLR model, as evidenced by the MAPE, RMS and MAE error values. These error values were 1.701, 0.565 and 0.425 for the MLP model and 6.458, 0.828 and 0.690 for the MLR model, respectively. The effectiveness of feed-forward neural networks in predicting the yield of oilseed rape and mustard grown in northeast India was also demonstrated by Kakati et al. [5]. The ANN model predicted crop yield for the Dhubri region with an RMS error of 11.3 and an R² value of 0.976, while the stepwise multiple linear regression (SMLR) model had an \mathbb{R}^2 value of 0.756 and an RMS error of 65.4 Ang et al. [76] investigated the feasibility of using different models, including MLR models and DNNs (deep neural networks), to predict the yield of oil palm grown in the state of Pahang (Malaysia). The deep neural network consisted of three hidden layers: the first layer contained 256 neurons, the second layer contained 480 neurons and the third layer contained 256 neurons. The model accurately predicted yield with an RMS error of 2.92 and an MAPE error of 0.09. In contrast, the MLR model had RMS and MAPE errors of 6.20 and 0.7, respectively. In addition, the R² coefficient of determination was 0.91 for the DNN model and 0.49 for the MLR model.

Other machine learning techniques are also effective in crop yield prediction, surpassing the quality of predictions made by classical MLR models. This is confirmed by the study by Sun et al. [77]. The authors used the random forest (RF) method to predict the yield of winter wheat grown in China. The study covered the years 2014–2018, and the integration of satellite, weather and geographical data was used to build the model. The average RMS error for the MLR model was 1229.97 and the coefficient of determination was 0.73. These results are significantly different from those obtained by the RF model. The mean RMS error for this model was more than 2.5 times smaller than that of the MLR model (465.32), and the R² coefficient was equal to 0.85. Zhao et al. [78] obtained similar results by investigating the ability of RF and MLR models in estimating the yield of winter wheat grown in the North China Plain. The researchers analyzed the applicability of these models at different periods of plant development. The results showed that the RF and MLR models obtained the best results for the period from the beginning of grain filling to the milk stage. However, the MLR model had a larger RMS error (778.0) and r-ratio (0.79) compared to the RF model, for which these parameters were 683.0 and 0.86, respectively.

A comparison of actual and predicted values (Figures 2 and 3) shows that the coefficient of determination for the N2 model ($R^2 = 0.8254$) was at a higher level than it was for the RS2 model ($R^2 = 0.5819$). These results show that the RS2 model had much weaker predictive properties with respect to the N2 model. However, multiple linear regression models, as already mentioned, are commonly used in yield prediction. This method has many limitations, such as the assumption of a linear relationship between the exploratory variable and the explanatory variable [17]. If the relationship between these variables is non-linear, the regression model will tend to perform poorly. In addition, linear regression assumes that the input variables are not correlated with each other. If there is multicollinearity in the dataset, then this assumption is violated, and the performance of the regression model will be reduced. Additionally, linear models are assumed to have a constant variance under error conditions (homoskedasticity), which is often not true. Another problem that hinders proper prediction using MLR models is the presence of outlier points, which significantly affect its performance [65]. On the other hand, ANN models, including those with MLP architecture, are capable of predicting agricultural crop yields even in the case of strong non-linear relationships between the independent variables and the dependent variable. In addition, the main function of neural networks is to identify hidden patterns and features in the dataset. This activity is made possible by the two most important parts of the network, i.e., the activation function and the weighting parameters [5]. From the research carried out, all the variables tested in the study are characterized by non-linear patterns. Therefore, the RS2 model could not properly estimate the yield. Our research shows that the choice of method for creating the model is a kind of compromise that requires its creator to have a very thorough knowledge of the test object. This ensures that accurate yield predictions are obtained. However, ANNs are also not free from certain limitations. One of the biggest is that neural network models require a lot of, sometimes very specific, input data to train [79]. Acquiring such data can often be cumbersome, and for regions where observational records are lacking, obtaining short-term predictions is significantly difficult [80]. In addition, the appropriate selection of independent characteristics must be supported by extensive knowledge of the issue being modeled [9]. In the present study, three categories of independent variables were used: weather data, agronomic information and phytophenological data. These variables are publicly available and the results of analyses involving these data are easy to interpret.

The inclusion of climatic conditions when modeling agronomic issues is an important element when seeking to obtain a high-performance model. The inclusion of information related to air temperature, sunshine and precipitation during the growing season is reasonable, as these factors strongly determine plant growth and development [81]. Plant productivity is significantly affected by the temperature distribution during the growing season. However, the influence of this factor is reduced when there is an adequate water supply to the plants [82]. This assumption was fulfilled in this work because data from typical years, without weather anomalies that affect the quality of the models, were selected for analysis. According to a study by Aubakirov et al. [19], the amount of precipitation and temperature had the greatest impact on the multiplicity of yield of wheat grown in the North Kazakhstan region. Consideration of these data by the authors made it possible to build a back-propagation artificial neural network model that predicted wheat yield multiplicity with an MAPE error of 12.02 and an RMS error of 3.368. Similar observations were reached by Nedbała et al. [59], who identified key meteorological factors affecting soybean (*Glycine max* [L.] Merrill) yield and harvest date. A sensitivity analysis of the MLP network found that the variable that most influenced soybean seed yield was air temperature in the second ten days of May. In contrast, the variables that most influenced soybean harvest date were rainfall totals in the first ten days of June and the first ten days of August. The inclusion of environmental variables in modeling is also highlighted by Vojnov et al. [60], who attribute significant effects on plant parameters and on the performance of ANN models to these data.

Many scientific disciplines and the agricultural industry commonly use phytophenological periods of plant development. Among other things, they are helpful in determining when to apply inputs. These phases have been standardized for ease of communication between agronomists, naturalists, breeders of new agricultural varieties, etc. [7]. In modeling agricultural crop yields, this information is exploited to enhance the efficiency of the models built. As reported by Shamsabadi et al. [20], the inclusion of data such as number of days to emergence, days to maturity and number of days to flowering in the model significantly affected the performance of the MLP model. The model predicted the seed yield of hybrid wheat that was grown in the northern part of Iran. In the present study, empirical data in the form of phytophenological periods were also used, which allowed for the construction of an N2 model with an MAPE error of 7.976 (Table 6) and a correlation coefficient of 0.936 (Table 3).

A sensitivity analysis of the N2 model showed that the independent variables with the greatest impact on pea seed yield under the conditions tested were the date of onset of maturity and the harvest date. These traits received a rank of 1 and 2, respectively. From Figures 4 and 6, we can observe that the later occurrence of onset of the maturity phase resulted in an increase in yield. The rate of transition of plants from one phenological phase to another depends on weather conditions [83]. Peas at the onset of maturity tolerate lower temperatures than those at the flowering stage. The length of this period is determined by average and minimum daily temperatures. Lower temperatures during this period favor the accumulation of starch in the seeds, thereby increasing yield [84]. Pea harvesting should be optimized based on weather conditions and seed moisture content. Figures 4 and 5 show that harvesting at a later date has a positive effect on yield. Pea varieties grown in Poland are characterized by uneven maturation. At the beginning, pods located at the lower part of the plant ripen, and pods located in the higher parts of the plant ripen at the end [85]. Harvesting too early may result in the upper pods not reaching the stage of technical maturity and not accumulating enough starch, proteins or other assimilates; thus, the weight of 1000 seeds may be lower than that of the seeds placed in the lower pods. It should be remembered, however, that harvesting peas from the field too late may result in a decrease in yield due to lodging of the plants and pod breakage [86].

The sum of precipitation and mean air temperature are the variables ranked 3 and 4 in the sensitivity analysis of the network. Weather conditions during the growing season of plants are one of the most important environmental factors affecting plant growth and development. Temperature and the amount of rainfall vary the yield of peas from one crop year to another [87]. A study conducted by Pandey et al. [88] proved that water deficiency in pea cultivation reduces the photosynthetic efficiency of plants, disrupts nutrient transport and affects structural changes in leaves due to the presence of reactive oxygen species.

15 of 19

These changes ultimately lead to a decrease in plant yield. Therefore, the optimum rainfall over the growing season of peas should be 280 mm on light soil, 250 mm on medium soil and 22 mm on compact soils [85].

The results obtained from the study show that the N2 model can be a promising information tool due to its accurate prediction of seed yield in pea. Our study further confirmed the hypothesis that the right approach to independent trait selection supports the process of identifying the most important variables affecting yield [76]. Such models may be of interest to breeders of new pea varieties. Knowing the variables with the greatest impact on yield, it is possible to improve new varieties by optimizing certain time intervals in the phenology so as to obtain a high final yield. An opportunity for the advancement of predictive models is the possibility of using new types of data, such as ground-based phenological imaging, or the use of the same dataset but of higher quality, such as high- or very high-resolution spectral data [7].

5. Conclusions

With climate change and an increasing global population, there is a growing need to better predict the yield of agricultural crops, as well as the correct way for farmers to grow their crops. The analyses conducted show that an artificial neural network model is a useful tool in predicting pea yield 20 days before harvest. The N2 model accurately predicted the independent variable with a correlation coefficient r > 0.9 and MAPE and RMS values of 7.976 and 0.443, respectively. At the same time, it was proven that the RS2 model is not able to accurately estimate pea yield. The model had an MAPE error of 148.585. Therefore, the potential practical application of this model in pea production is not possible. The choice of modeling technique is crucial in accurately estimating yield. Furthermore, modeling pea yields a few weeks before harvest carries promising possibilities for practical application.

Pre-harvest yield forecasting is a valuable source of information that is particularly relevant for farmers, agronomists and decision makers. Further research will focus on comparative analysis of the ANN model against other machine learning techniques such as RBF.

Author Contributions: Conceptualization, P.H., M.P. and G.N.; methodology, P.H., M.P. and G.N.; validation, M.P. and G.N.; formal analysis, M.P.; investigation, P.H.; resources, P.H.; data curation, P.H.; writing—original draft preparation, P.H.; writing—review and editing, P.H., M.P. and G.N.; supervision, G.N.; project administration, M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable. Data Availability Statement The data presented in this study are available on request from the corresponding author.

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

Abbreviations

ANN—artificial neural networks; COBORU—Research Center for Cultivar Testing; CNN—convolutional neural networks; DNN—deep neural network; FLOWE—number of days from 1 January to the beginning of flowering; GEN—general variety of peas; HAR—number of days from 1 January to the date of harvesting; INI_MA—number of days from 1 January to onset of maturity; kNN—k nearest neighbors; K20_C—K₂O content in the soil; K₂O—potassium oxide; K2O_F—Total potassium from mineral fertilizers; kg—kilogram; MAE—mean absolute error; MAPE—mean absolute percentage error; MAX—maximum error determined for the whole model; MAXP—maximum percentage error; MgO—magnesium oxide; MGO_C—MgO content in the soil; ML—machine learning; MLP—multilayer perceptron; MLR—multiple linear regression; n—number of observations; N_F—total nitrogen from mineral fertilizers; N2—built its own neural network model; P_EMER—number of days from 1 January to the beginning of plant emergence; P_HIG—plant height; P₂O₅—phosphorus(V) oxide; P₂O₅_C—P₂O₅ content in the soil; P₂O₅_F—total phosphorus from mineral fertilizers; PH—Soil pH; PROT—Percentage of protein in pea seeds; RAE—global relative error of model approximation; RAIN—total rainfall from sowing date to July 14; RF—random forest; RMS—root mean square error; RS2—built its own linear regression model; SOWI—number of days from 1 January to sowing date; SUN—total sunshine from sowing date to 14 July; SVM—support vector machines; TECH_M—number of days from 1 January to technical maturity; TEMP—average air temperature from sowing date to July 14; WEGW—number of plant growing days; y'_i —predictive values, obtained with the model; y_i —actual values.

References

- Szparaga, A.; Kuboń, M.; Kocira, S.; Czerwińska, E.; Pawłowska, A.; Hara, P.; Kobus, Z.; Kwaśniewski, D. Towards sustainable agriculture-agronomic and economic effects of biostimulant use in common bean cultivation. *Sustainability* 2019, *11*, 4575. [CrossRef]
- Rokhafrouz, M.; Latifi, H.; Abkar, A.A.; Wojciechowski, T.; Czechlowski, M.; Naieni, A.S.; Maghsoudi, Y.; Niedbała, G. Simplified and Hybrid Remote Sensing-Based Delineation of Management Zones for Nitrogen Variable Rate Application in Wheat. *Agriculture* 2021, 11, 1104. [CrossRef]
- Kukal, M.S.; Irmak, S. Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. Sci. Rep. 2018, 8, 3450. [CrossRef] [PubMed]
- Khavse, R.; Singh, R.; Manikandan, N.; Chaudhary, J. Influence of Temperature on Rapeseed-Mustard Yield at Selected Locations in Chhattisgarh State. *Curr. World Environ.* 2014, *9*, 1034–1036. [CrossRef]
- Kakati, N.; Deka, R.L.; Das, P.; Goswami, J.; Khanikar, P.G.; Saikia, H. Forecasting yield of rapeseed and mustard using multiple linear regression and ANN techniques in the Brahmaputra valley of Assam, North East India. *Theor. Appl. Climatol.* 2022, 150, 1201–1215. [CrossRef]
- Chergui, N.; Kechadi, M.-T.; McDonnell, M. The Impact of Data Analytics in Digital Agriculture: A Review. In Proceedings of the 2020 International Multi-Conference on: "Organization of Knowledge and Advanced Technologies" (OCTA), Tunis, Tunisia, 6–8 February 2020; IEEE: New York, NY, USA, 2020; pp. 1–13.
- Niedbała, G.; Kurek, J.; Świderski, B.; Wojciechowski, T.; Antoniuk, I.; Bobran, K. Prediction of Blueberry (*Vaccinium corymbosum* L.) Yield Based on Artificial Intelligence Methods. *Agriculture* 2022, 12, 2089. [CrossRef]
- 8. He, L.; Fang, W.; Zhao, G.; Wu, Z.; Fu, L.; Li, R.; Majeed, Y.; Dhupia, J. Fruit yield prediction and estimation in orchards: A state-of-the-art comprehensive review for both direct and indirect methods. *Comput. Electron. Agric.* 2022, 195, 106812. [CrossRef]
- 9. Hara, P.; Piekutowska, M.; Niedbała, G. Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data. *Land* **2021**, *10*, 609. [CrossRef]
- 10. Yildirim, T.; Moriasi, D.N.; Starks, P.J.; Chakraborty, D. Using Artificial Neural Network (ANN) for Short-Range Prediction of Cotton Yield in Data-Scarce Regions. *Agronomy* **2022**, *12*, 828. [CrossRef]
- 11. Ali, A.; Rondelli, V.; Martelli, R.; Falsone, G.; Lupia, F.; Barbanti, L. Management Zones Delineation through Clustering Techniques Based on Soils Traits, NDVI Data, and Multiple Year Crop Yields. *Agriculture* **2022**, *12*, 231. [CrossRef]
- 12. Wang, J.; Si, H.; Gao, Z.; Shi, L. Winter Wheat Yield Prediction Using an LSTM Model from MODIS LAI Products. *Agriculture* 2022, 12, 1707. [CrossRef]
- 13. Johnson, M.D.; Hsieh, W.W.; Cannon, A.J.; Davidson, A.; Bédard, F. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agric. For. Meteorol.* **2016**, *218–219*, 74–84. [CrossRef]
- 14. Gonzalez-Sanchez, A.; Frausto-Solis, J.; Ojeda-Bustamante, W. Attribute Selection Impact on Linear and Nonlinear Regression Models for Crop Yield Prediction. *Sci. World J.* **2014**, 2014, 509429. [CrossRef] [PubMed]
- Peng, J.; Kim, M.; Kim, Y.; Jo, M.; Kim, B.; Sung, K.; Lv, S. Constructing Italian ryegrass yield prediction model based on climatic data by locations in South Korea. *Grassl. Sci.* 2017, 63, 184–195. [CrossRef]
- 16. Niedbala, G.; Kozlowski, R.J. Application of artificial neural networks for multi-criteria yield prediction of winter wheat. *J. Agric. Sci. Technol.* **2019**, *21*, 51–61.
- Piekutowska, M.; Niedbała, G.; Piskier, T.; Lenartowicz, T.; Pilarski, K.; Wojciechowski, T.; Pilarska, A.A.; Czechowska-Kosacka, A. The Application of Multiple Linear Regression and Artificial Neural Network Models for Yield Prediction of Very Early Potato Cultivars before Harvest. *Agronomy* 2021, 11, 885. [CrossRef]
- Niedbała, G.; Wróbel, B.; Piekutowska, M.; Zielewicz, W.; Paszkiewicz-Jasińska, A.; Wojciechowski, T.; Niazian, M. Application of Artificial Neural Networks Sensitivity Analysis for the Pre-Identification of Highly Significant Factors Influencing the Yield and Digestibility of Grassland Sward in the Climatic Conditions of Central Poland. *Agronomy* 2022, *12*, 1133. [CrossRef]
- 19. Aubakirova, G.; Ivel, V.; Gerassimova, Y.; Moldakhmetov, S.; Petrov, P. Application of artificial neural network for wheat yield forecasting. *East. Eur. J. Enterp. Technol.* **2022**, *3*, 31–39. [CrossRef]
- Shamsabadi, E.E.; Sabouri, H.; Soughi, H.; Sajadi, S.J. Using of Molecular Markers in Prediction of Wheat (*Triticum aestivum* L.) Hybrid Grain Yield Based on Artificial Intelligence Methods and Multivariate Statistics. *Russ. J. Genet.* 2022, 58, 603–611. [CrossRef]

- 21. Khaki, S.; Wang, L.; Archontoulis, S.V. A CNN-RNN Framework for Crop Yield Prediction. *Front. Plant Sci.* 2020, 10, 1750. [CrossRef]
- 22. Sabatino, L.; D'Anna, F.; Iapichino, G.; Moncada, A.; D'Anna, E.; De Pasquale, C. Interactive Effects of Genotype and Molybdenum Supply on Yield and Overall Fruit Quality of Tomato. *Front. Plant Sci.* **2019**, *9*, 1922. [CrossRef] [PubMed]
- Awad, M. Toward Precision in Crop Yield Estimation Using Remote Sensing and Optimization Techniques. Agriculture 2019, 9, 54. [CrossRef]
- Nazir, A.; Ullah, S.; Saqib, Z.A.; Abbas, A.; Ali, A.; Iqbal, M.S.; Hussain, K.; Shakir, M.; Shah, M.; Butt, M.U. Estimation and Forecasting of Rice Yield Using Phenology-Based Algorithm and Linear Regression Model on Sentinel-II Satellite Data. *Agriculture* 2021, 11, 1026. [CrossRef]
- 25. Feizi, H.; Sattari, M.T.; Prasad, R.; Apaydin, H. Comparative analysis of deep and machine learning approaches for daily carbon monoxide pollutant concentration estimation. *Int. J. Environ. Sci. Technol.* **2023**, *20*, 1753–1768. [CrossRef]
- 26. Meerasri, J.; Sothornvit, R. Artificial neural networks (ANNs) and multiple linear regression (MLR) for prediction of moisture content for coated pineapple cubes. *Case Stud. Therm. Eng.* **2022**, *33*, 101942. [CrossRef]
- Ge, J.; Zhao, L.; Yu, Z.; Liu, H.; Zhang, L.; Gong, X.; Sun, H. Prediction of Greenhouse Tomato Crop Evapotranspiration Using XGBoost Machine Learning Model. *Plants* 2022, 11, 1923. [CrossRef]
- Niedbała, G. Application of artificial neural networks for multi-criteria yield prediction of winter rapeseed. Sustainability 2019, 11, 533. [CrossRef]
- Wojciechowski, T.; Niedbala, G.; Czechlowski, M.; Nawrocka, J.R.; Piechnik, L.; Niemann, J. Rapeseed seeds quality classification with usage of VIS-NIR fiber optic probe and artificial neural networks. In Proceedings of the 2016 International Conference on Optoelectronics and Image Processing (ICOIP), Warsaw, Poland, 10–12 June 2016; IEEE: Warsaw, Poland, 2016; pp. 44–48.
- 30. van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* 2020, 177, 105709. [CrossRef]
- 31. Ballesteros, R.; Intrigliolo, D.S.; Ortega, J.F.; Ramírez-Cuesta, J.M.; Buesa, I.; Moreno, M.A. Vineyard yield estimation by combining remote sensing, computer vision and artificial neural network techniques. *Precis. Agric.* 2020, *21*, 1242–1262. [CrossRef]
- Crusiol, L.G.T.; Sun, L.; Sibaldelli, R.N.R.; Junior, V.F.; Furlaneti, W.X.; Chen, R.; Sun, Z.; Wuyun, D.; Chen, Z.; Nanni, M.R.; et al. Strategies for monitoring within-field soybean yield using Sentinel-2 Vis-NIR-SWIR spectral bands and machine learning regression methods. *Precis. Agric.* 2022, 23, 1093–1123. [CrossRef]
- Niedbała, G.; Kurasiak-Popowska, D.; Stuper-Szablewska, K.; Nawracała, J. Application of Artificial Neural Networks to Analyze the Concentration of Ferulic Acid, Deoxynivalenol, and Nivalenol in Winter Wheat Grain. Agriculture 2020, 10, 127. [CrossRef]
- 34. Chergui, N. Durum wheat yield forecasting using machine learning. Artif. Intell. Agric. 2022, 6, 156–166. [CrossRef]
- 35. Phan, P.; Chen, N.; Xu, L.; Dao, D.M.; Dang, D. NDVI Variation and Yield Prediction in Growing Season: A Case Study with Tea in Tanuyen Vietnam. *Atmosphere* **2021**, *12*, 962. [CrossRef]
- Bouras, E.H.; Jarlan, L.; Er-Raki, S.; Balaghi, R.; Amazirh, A.; Richard, B.; Khabba, S. Cereal Yield Forecasting with Satellite Drought-Based Indices, Weather Data and Regional Climate Indices Using Machine Learning in Morocco. *Remote Sens.* 2021, 13, 3101. [CrossRef]
- 37. Taşan, S.; Cemek, B.; Taşan, M.; Cantürk, A. Estimation of eggplant yield with machine learning methods using spectral vegetation indices. *Comput. Electron. Agric.* 2022, 202, 107367. [CrossRef]
- Jeevaganesh, R.; Harish, D.; Priya, B. A Machine Learning-based Approach for Crop Yield Prediction and Fertilizer Recommendation. In Proceedings of the 2022 6th International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 28–30 April 2022; IEEE: New York, NY, USA, 2022; pp. 1330–1334.
- 39. Tugrul, B.; Elfatimi, E.; Eryigit, R. Convolutional Neural Networks in Detection of Plant Leaf Diseases: A Review. *Agriculture* **2022**, *12*, 1192. [CrossRef]
- Dayal, M.; Gupta, M.; Gupta, M.; Bara, A.R.; Chaubey, C. Introduction to Machine Learning Methods With Application in Agriculture. In *Applying Drone Technologies and Robotics for Agricultural Sustainability*; IGI Global: Hershey, PA, USA, 2023; pp. 184–203.
- Dhillon, M.S.; Dahms, T.; Kuebert-Flock, C.; Rummler, T.; Arnault, J.; Steffan-Dewenter, I.; Ullmann, T. Integrating random forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape. *Front. Remote Sens.* 2023, *3*, 1010978. [CrossRef]
- 42. Khalifani, S.; Darvishzadeh, R.; Azad, N.; Seyed Rahmani, R. Prediction of sunflower grain yield under normal and salinity stress by RBF, MLP and, CNN models. *Ind. Crops Prod.* **2022**, *189*, 115762. [CrossRef]
- Maya Gopal, P.S.; Bhargavi, R. A novel approach for efficient crop yield prediction. *Comput. Electron. Agric.* 2019, 165, 104968. [CrossRef]
- 44. Tadeusiewicz, R. Elementarne Wprowadzenie Do Techniki Sieci Neuronowych z Przykładowymi Programami; Akademicka Oficyna Wydawnicza PLJ: Warsaw, Poland, 1998.
- 45. Li, X.; Hu, T.; Gong, P.; Du, S.; Chen, B.; Li, X.; Dai, Q. Mapping Essential Urban Land Use Categories in Beijing with a Fast Area of Interest (AOI)-Based Method. *Remote Sens.* **2021**, *13*, 477. [CrossRef]
- 46. Sabzi-Nojadeh, M.; Niedbała, G.; Younessi-Hamzekhanlu, M.; Aharizad, S.; Esmaeilpour, M.; Abdipour, M.; Kujawa, S.; Niazian, M. Modeling the Essential Oil and Trans-Anethole Yield of Fennel (*Foeniculum vulgare* Mill. var. vulgare) by Application Artificial Neural Network and Multiple Linear Regression Methods. *Agriculture* 2021, 11, 1191. [CrossRef]

- Abrosimov, M.; Brovko, A. High Generalization Capability Artificial Neural Network Architecture Based on RBF-Network. In Proceedings of the ICIT 2019: Recent Research in Control Engineering and Decision Making, Saratov, Russia, 7–8 February 2019; Springer: Cham, Switzerland, 2019; pp. 67–78.
- 48. Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* 2018, 18, 2674. [CrossRef] [PubMed]
- Bhojani, S.H.; Bhatt, N. Wheat crop yield prediction using new activation functions in neural network. *Neural Comput. Appl.* 2020, 32, 13941–13951. [CrossRef]
- Bazrafshan, O.; Ehteram, M.; Dashti Latif, S.; Feng Huang, Y.; Yenn Teo, F.; Najah Ahmed, A.; El-Shafie, A. Predicting crop yields using a new robust Bayesian averaging model based on multiple hybrid ANFIS and MLP models. *Ain Shams Eng. J.* 2022, 13, 101724. [CrossRef]
- 51. Torsoni, G.B.; de Oliveira Aparecido, L.E.; dos Santos, G.M.; Chiquitto, A.G.; da Silva Cabral Moraes, J.R.; de Souza Rolim, G. Soybean yield prediction by machine learning and climate. *Theor. Appl. Climatol.* **2023**, *151*, 1709–1725. [CrossRef]
- 52. Soroush, F.; Ehteram, M.; Seifi, A. Uncertainty and spatial analysis in wheat yield prediction based on robust inclusive multiple models. *Environ. Sci. Pollut. Res.* 2022, *30*, 20887–20906. [CrossRef] [PubMed]
- Heng, S.Y.; Ridwan, W.M.; Kumar, P.; Ahmed, A.N.; Fai, C.M.; Birima, A.H.; El-Shafie, A. Artificial neural network model with different backpropagation algorithms and meteorological data for solar radiation prediction. *Sci. Rep.* 2022, *12*, 10457. [CrossRef]
- 54. Hara, P.; Piekutowska, M.; Niedbała, G. Prediction of Protein Content in Pea (*Pisum sativum* L.) Seeds Using Artificial Neural Networks. *Agriculture* 2022, 13, 29. [CrossRef]
- 55. Research Centre for Cultivar Testing (COBORU). Available online: https://coboru.gov.pl/ (accessed on 20 October 2022).
- Niedbała, G.; Tratwal, A.; Piekutowska, M.; Wojciechowski, T.; Uglis, J. A Framework for Financing Post-Registration Variety Testing System: A Case Study from Poland. *Agronomy* 2022, 12, 325. [CrossRef]
- 57. Zintegrowana Platforma Edukacyjna. Available online: https://zpe.gov.pl/a/cechy-klimatu-polski/DbdxuNlhI (accessed on 10 January 2023).
- 58. Wiatr, K. Rośliny rolnicze. In *Metodyka Badania Wartości Gospodarczej Odmian (WGO) Roślin Uprawnych;* Centralny Ośrodek Badania Odmian Roślin Uprawnych: Słupia Wielka, Poland, 1998.
- Niedbała, G.; Kurasiak-Popowska, D.; Piekutowska, M.; Wojciechowski, T.; Kwiatek, M.; Nawracała, J. Application of Artificial Neural Network Sensitivity Analysis to Identify Key Determinants of Harvesting Date and Yield of Soybean (*Glycine max* [L.] Merrill) Cultivar Augusta. *Agriculture* 2022, 12, 754. [CrossRef]
- Vojnov, B.; Jaćimović, G.; Šeremešić, S.; Pezo, L.; Lončar, B.; Krstić, Đ.; Vujić, S.; Ćupina, B. The Effects of Winter Cover Crops on Maize Yield and Crop Performance in Semiarid Conditions—Artificial Neural Network Approach. *Agronomy* 2022, 12, 2670. [CrossRef]
- 61. Geetha, M.C.S.; Elizabeth Shanthi, I. Forecasting the Crop Yield Production in Trichy District Using Fuzzy C-Means Algorithm and Multilayer Perceptron (MLP). *Int. J. Knowl. Syst. Sci.* 2020, 11, 83–98. [CrossRef]
- Pentoś, K.; Mbah, J.T.; Pieczarka, K.; Niedbała, G.; Wojciechowski, T. Evaluation of Multiple Linear Regression and Machine Learning Approaches to Predict Soil Compaction and Shear Stress Based on Electrical Parameters. *Appl. Sci.* 2022, 12, 8791. [CrossRef]
- 63. Gorzelany, J.; Belcar, J.; Kuźniar, P.; Niedbała, G.; Pentoś, K. Modelling of Mechanical Properties of Fresh and Stored Fruit of Large Cranberry Using Multiple Linear Regression and Machine Learning. *Agriculture* **2022**, *12*, 200. [CrossRef]
- 64. Shankar, T.; Malik, G.C.; Banerjee, M.; Dutta, S.; Praharaj, S.; Lalichetti, S.; Mohanty, S.; Bhattacharyay, D.; Maitra, S.; Gaber, A.; et al. Prediction of the Effect of Nutrients on Plant Parameters of Rice by Artificial Neural Network. *Agronomy* **2022**, *12*, 2123. [CrossRef]
- 65. Khan, S.N.; Li, D.; Maimaitijiang, M. A Geographically Weighted Random Forest Approach to Predict Corn Yield in the US Corn Belt. *Remote Sens.* **2022**, *14*, 2843. [CrossRef]
- Arroyo, Á.; Cambra, C.; Basurto, N.; Rad, C.; Navarro, M.; Herrero, Á. Regression Techniques to Predict the Growth of Potato Tubers. In Proceedings of the 17th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2022), Salamanca, Spain, 5–7 September 2022; Springer: Cham, Switzerland, 2023; pp. 217–225.
- 67. Tadeusiewicz, R.; Szaleniec, M. Leksykon Sieci Neuronowych; Fundacja na Rzecz Promocji Nauki Polskiej: Wrocław, Poland, 2015.
- 68. Zhang, Q.; Wang, K.; Han, Y.; Liu, Z.; Yang, F.; Wang, S.; Zhao, X.; Zhao, C. A crop variety yield prediction system based on variety yield data compensation. *Comput. Electron. Agric.* **2022**, 203, 107460. [CrossRef]
- 69. Piekutowska, M.; Adamski, M.; Czechowska-Kosacka, A.; Wójcik Oliveira, K.; Niedbała, G.; Wojciechowski, T.; Czechlowski, M. Modeling methods of predicting potato yield—Examples and possibilities of application. *J. Res. Appl. Agric. Eng.* **2018**, *63*, 176.
- 70. Elbeltagi, A.; Zhang, L.; Deng, J.; Juma, A.; Wang, K. Modeling monthly crop coefficients of maize based on limited meteorological data: A case study in Nile Delta, Egypt. *Comput. Electron. Agric.* **2020**, *173*, 105368. [CrossRef]
- Gené-Mola, J.; Gregorio, E.; Auat Cheein, F.; Guevara, J.; Llorens, J.; Sanz-Cortiella, R.; Escolà, A.; Rosell-Polo, J.R. Fruit detection, yield prediction and canopy geometric characterization using LiDAR with forced air flow. *Comput. Electron. Agric.* 2020, 168, 105121. [CrossRef]
- 72. Ronchetti, G.; Manfron, G.; Weissteiner, C.J.; Seguini, L.; Nisini Scacchiafichi, L.; Panarello, L.; Baruth, B. Remote sensing crop group-specific indicators to support regional yield forecasting in Europe. *Comput. Electron. Agric.* 2023, 205, 107633. [CrossRef]
- 73. Atamanyuk, I.; Havrysh, V.; Nitsenko, V.; Diachenko, O.; Tepliuk, M.; Chebakova, T.; Trofimova, H. Forecasting of Winter Wheat Yield: A Mathematical Model and Field Experiments. *Agriculture* **2022**, *13*, 41. [CrossRef]

- Kumari, P.; Mishra, G.C.; Srivastava, C.P. Statistical models for forecasting pigeonpea yield in Varanasi region. *J. Agrometeorol.* 2016, 18, 306–310. [CrossRef]
- 75. Sousa, A.M.d.C.B.d.; Silva, V.B.d.; Lopes, Â.C.d.A.; Gomes, R.L.F.; Carvalho, L.C.B. Prediction of grain yield, adaptability, and stability in landrace varieties of lima bean (*Phaseolus lunatus* L.). *Crop Breed. Appl. Biotechnol.* **2020**, *20*, 1–7. [CrossRef]
- 76. Ang, Y.; Shafri, H.Z.M.; Lee, Y.P.; Bakar, S.A.; Abidin, H.; Mohd Junaidi, M.U.U.; Hashim, S.J.; Che'Ya, N.N.; Hassan, M.R.; Lim, H.S.; et al. Oil palm yield prediction across blocks from multi-source data using machine learning and deep learning. *Earth Sci. Inform.* 2022, 15, 2349–2367. [CrossRef]
- 77. Sun, Y.; Zhang, S.; Tao, F.; Aboelenein, R.; Amer, A. Improving Winter Wheat Yield Forecasting Based on Multi-Source Data and Machine Learning. *Agriculture* **2022**, 12, 571. [CrossRef]
- 78. Zhao, Y.; Xiao, D.; Bai, H.; Tang, J.; Liu, D.L.; Qi, Y.; Shen, Y. The Prediction of Wheat Yield in the North China Plain by Coupling Crop Model with Machine Learning Algorithms. *Agriculture* **2022**, *13*, 99. [CrossRef]
- 79. Li, Y.; Guan, K.; Yu, A.; Peng, B.; Zhao, L.; Li, B.; Peng, J. Toward building a transparent statistical model for improving crop yield prediction: Modeling rainfed corn in the U.S. *Field Crops Res.* **2019**, *234*, 55–65. [CrossRef]
- 80. Filippi, P.; Whelan, B.M.; Vervoort, R.W.; Bishop, T.F.A. Mid-season empirical cotton yield forecasts at fine resolutions using large yield mapping datasets and diverse spatial covariates. *Agric. Syst.* **2020**, *184*, 102894. [CrossRef]
- Ruan, G.; Li, X.; Yuan, F.; Cammarano, D.; Ata-UI-Karim, S.T.; Liu, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cao, Q. Improving wheat yield prediction integrating proximal sensing and weather data with machine learning. *Comput. Electron. Agric.* 2022, 195, 106852. [CrossRef]
- 82. Skrzyczyńska, J.; Gąsiorowska, B. Uprawa Roślin; UPW: Wrocław, Poland, 2020; pp. 49-210.
- Lamichaney, A.; Parihar, A.K.; Hazra, K.K.; Dixit, G.P.; Katiyar, P.K.; Singh, D.; Singh, A.K.; Kumar, N.; Singh, N.P. Untangling the Influence of Heat Stress on Crop Phenology, Seed Set, Seed Weight, and Germination in Field Pea (*Pisum sativum* L.). Front. Plant Sci. 2021, 12, 635868. [CrossRef] [PubMed]
- 84. Grzebisz, W. Technologia Nawożenia Roślin Uprawnych—Fizjologia Plonowania; Tom 1 Olei.; Powszechne Wydawnictwo Rolnicze i Lesne: Poznań, Poland, 2021.
- 85. Kotecki, A. Uprawa Roślin Tom III.; Wydawnictwo Uniwersytetu Przyrodniczego we Wrocławiu: Wrocław, Poland, 2020.
- 86. Singh, A.K.; Srivastava, C.P. Effect of plant types on grain yield and lodging resistance in pea (*Pisum sativum* L.). *Indian J. Genet. Plant Breed.* **2015**, *75*, 69. [CrossRef]
- 87. Wysokinski, A.; Lozak, I. The Dynamic of Nitrogen Uptake from Different Sources by Pea (*Pisum sativum* L.). Agriculture 2021, 11, 81. [CrossRef]
- Pandey, J.; Devadasu, E.; Saini, D.; Dhokne, K.; Marriboina, S.; Raghavendra, A.S.; Subramanyam, R. Reversible changes in structure and function of photosynthetic apparatus of pea (*Pisum sativum*) leaves under drought stress. *Plant J.* 2023, 113, 60–74. [CrossRef] [PubMed]

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.