Article

Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data

Patryk Hara [1],*[ID], Magdalena Piekutowska [2][ID] and Gniewko Niedbała [3][ID]

[1] Department of Agrobiotechnology, Koszalin University of Technology, Racławicka 15–17, 75-620 Koszalin, Poland
[2] Department of Geoecology and Geoinformation, Institute of Biology and Earth Sciences, Pomeranian University in Słupsk, 27 Partyzantów St., 76-200 Słupsk, Poland; magdalena.piekutowska@apsl.edu.pl
[3] Department of Biosystems Engineering, Faculty of Environmental and Mechanical Engineering, Poznań University of Life Sciences, Wojska Polskiego 50, 60-627 Poznań, Poland; gniewko.niedbala@up.poznan.pl
* Correspondence: patryk.hara@s.tu.koszalin.pl

Abstract: Knowing the expected crop yield in the current growing season provides valuable information for farmers, policy makers, and food processing plants. One of the main benefits of using reliable forecasting tools is generating more income from grown crops. Information on the amount of crop yielding before harvesting helps to guide the adoption of an appropriate strategy for managing agricultural products. The difficulty in creating forecasting models is related to the appropriate selection of independent variables. Their proper selection requires a perfect knowledge of the research object. The following article presents and discusses the most commonly used independent variables in agricultural crop yield prediction modeling based on artificial neural networks (ANNs). Particular attention is paid to environmental variables, such as climatic data, air temperature, total precipitation, insolation, and soil parameters. The possibility of using plant productivity indices and vegetation indices, which are valuable predictors obtained due to the application of remote sensing techniques, are analyzed in detail. The paper emphasizes that the increasingly common use of remote sensing and photogrammetric tools enables the development of precision agriculture. In addition, some limitations in the application of certain input variables are specified, as well as further possibilities for the development of non-linear modeling, using artificial neural networks as a tool supporting the practical use of and improvement in precision farming techniques.

Keywords: crop yield prediction; independent variables; ANN; remote sensing

1. Introduction

Maximizing yield while minimizing costs and caring for the environment are the basic goals of agricultural production [1]. Early detection and management of problems limiting plant production can help to increase yields and thus obtain more income for the farm [2].

Some of the techniques used for forecasting the yield of crops are statistical models and machine learning algorithms, which enable yield estimation during the growing season [3–5]. Knowing the predicted crop size in a specific year can be helpful in making decisions concerning the seasonal planning of cultivation or storage areas [6]. On the basis of the yield forecast, it is possible to improve the profitability of a farm, as well as to balance the amount of production means used, such as fertilizers [7] or plant protection products. The balanced consumption of these products leads to both a reduction in energy inputs on the farm and in human labor inputs. Ultimately, the farming company experiences an increase in income due to the lower production costs [8]. In addition, an early and reliable monitoring of crop yields can help policymakers and grain marketing agencies to plan

imports and exports. Yield forecasting increases the profit of processing plants; forecasting tools provide valuable information on the basis of which it is possible to estimate the amount of raw material that these companies will be able to process [9–11].

Crop prediction for the current growing season is a relatively difficult task due to the complexity of the relationship between the plant growth process and environmental factors such as weather or soil variability [12]. As such, the final yielding result is determined by many independent variables. Additionally, in yield prediction, possible interactions between factors or groups of yield-forming factors should be considered [13,14]. The application of statistical modeling methods such as multiple linear regression (MLR) does not always produce satisfactory results as they assume a linear relationship between the variables related to plant production, whereas these relationships usually follow a nonlinear course [15–17]. Failure by regression models to meet the assumption of multiple collinearity of dependent and independent variables results in poor to moderate results in the prediction of crop yield [6,18].

An essential element in predictive modeling is the accurate assessment of the correctness of the model's functionality. For this purpose, ex post forecast quality indices are used. One of the most frequently used forecast error indicators is the mean absolute percentage error (MAPE), which is calculated according to Formula (1) [19–25]. MAPE measures the error in percent and specifies the average percentage deviation between the forecast value and the actual implementation [26]. Peng et al. [27] reported that if the MAPE value is below 10%, the degree of model goodness-of-fit is perfect; if it falls within the range of 10–20%, the model goodness-of-fit is good. In the range of 20% to 30%, the error level is acceptable, while error exceeding 30% is considered a poor result and the model should therefore be rejected. According to Kim and Kim [28], when the real value is close to or equal to zero, the MAPE provides infinite or undefined values, which is considered to be its significant disadvantage. Therefore, to verify in detail the assessment of the efficiency of the prediction model, a combination of the MAPE and the root mean square error (RMSE) is used [29]. RMSE indicates the absolute fit of the observed data point to the predicted values. It is determined, according to Formula (2), as the root of the second order from the mean square of all errors [26,30]. The lower the value of MAPE and RMSE, the higher the accuracy of the obtained predictive model [31]. Other measures to determine error include mean absolute error (MAE), mean squared error (MSE), relative absolute error (RAE), root relative squared error (RRSE), and others [32–35].

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_i - y\prime_i}{y_i} \right| \times 100 \tag{1}$$

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} (y_i - y\prime_i)^2}, \tag{2}$$

where y_i represents the actual yield, $y\prime_i$ is the predicted yield, and n is the sample size.

ANNs function similar to biological neural systems in terms of the ability to learn and acquire resistance to errors [36]. In artificial neural networks, each neuron is assigned an appropriate value of the coefficient (weight), which determines its properties and role in the process of solving a given problem by the network. The weight values are changed during the learning process to reduce the error between the result given by the teacher and that obtained by the network. The set of weight values established in all neurons determines the knowledge of the neural network [37]. Machine learning methods, including artificial neural networks, are characterized by a high self-adaptation ability [38], which enables their application to address many scientific issues [39]. ANNs are mainly used to solve regression and classification problems [40]. ANNs have been used, among others, for the prediction of long-term climatic data [41], in the construction industry [42], and in medicine to identify neoplastic lesions based on mammogram images [43]. In addition, ANNs have been used in the power industry for the estimation of demand for electricity [44] and predicting the power of photovoltaic panels [45], as well as water quality forecasting in

agricultural drainage river basins [46] or predicting bioethanol production process from lignocellulosic biomass [47].

The application of artificial neural networks (ANNs) in agriculture solved the problem of the lack of linearity between the crop yield and independent variables. An essential feature of ANNs is the ability to learn by means of two variants of learning: supervised and unsupervised learning. The supervised learning process is based on the training set, which includes the learning cases along with the model answers provided to the model. This allows the network answers to be matched with the pattern answers. Training a neural network makes it capable of solving a task similar to the one on which it was trained. Unsupervised learning is based only on providing a series of sample inputs, without considering any information about the expected outputs. An accurately designed neural network can only use the observations of input signals and, on this basis, build an algorithm for its operation [40,48]. The ability to transfer the trained knowledge to new cases is known as generalization. Overfitting in generalization is a risk, which causes excessive fitting to irrelevant learning cases. Overfitting of the network results in poor generalization [49]. According to Caselli et al. [50], artificial neural networks are one of the best tools for obtaining information from imprecise and non-linear data. An additional advantage of artificial neural networks is the possibility of using qualitative (linguistic) variables without the need to code them first, as is the case with conventional statistical techniques [51]. One of the widely used comparative methods for ANN-based analyses is multiple linear regression (MLR) [52–57]. Many studies have demonstrated the advantage of ANNs over MLR in forecasting crop yields. Zaefizadeh et al. [58] analyzed the possibility of using ANN and MLR to forecast the yield of barley grown in Ardabil, Iran. In the study, a multilayer perceptron (MLP) with three input neurons, 15 neurons in the hidden layer, and one output neuron was applied. Basing on the mean absolute error, the authors discovered that ANN was more accurate than MLR. The obtained error values were 0.21 and 0.22 t·ha^{-1}, respectively. Similar results were also reported by Niazian et al. [59] in forecasting the yield size of ajowan seeds grown in Iran. The authors obtained higher accuracy using the ANN model prediction than with MLR. The ANN model used a network with the SigmoidAxon transfer function and the Levenberg–Marquart learning algorithm. The RMSE of the ANN model was 0.15 t·ha^{-1}, and the coefficient of determination (R^2) was 0.93. For the MLR model, the RMSE was 0.21 t·ha^{-1} and the R^2 was 0.79. Artificial neural networks were also compared with other crop yield estimation models. For example, Drummond et al. [15] compared the effectiveness of neural networks with projection pursuit regression and stepwise multiple linear regression. The analysis included fields located in the state of Missouri, USA. The study period covered 10 years and involved soybean and corn crops. The results showed that neural networks outperformed other techniques in terms of grain yield prediction quality, with R^2 values ranging from 0.31 to 0.74. Khaki et al. [60] analyzed the effectiveness of a hybrid model based on convolutional neural networks (CNNs) and recurrent neural networks. The developed model was compared with other popular methods such as random forest, deep fully connected neural networks, and LASSO. The models were used to forecast corn and soybean yields across the corn belt in the USA from 2016 to 2018. The model developed by the authors had a validation correlation coefficient ranging between 85.82% and 88.24%, and a training RMSE of 11.48 and 13.26. Jiang et al. [61] developed a long short-term memory model to forecast corn grain yields. The study area included nine states: Illinois, Indiana, Iowa, Kentucky, Michigan, Minnesota, Missouri, Wisconsin, and Ohio. The long short-term memory model accounted for 76% of yield variations and outperformed models like LASSO and random forest. Bornn and Zidek [62] used a Bayesian model to predict wheat grain yield. The crops located in the Canadian prairies were included in the analysis and historical data ranged from 1976–2006. The model obtained by the authors had high prediction quality with an RMSE of 5.35 and R^2 of 0.70. Figure 1 shows the steps of working with predictive models.
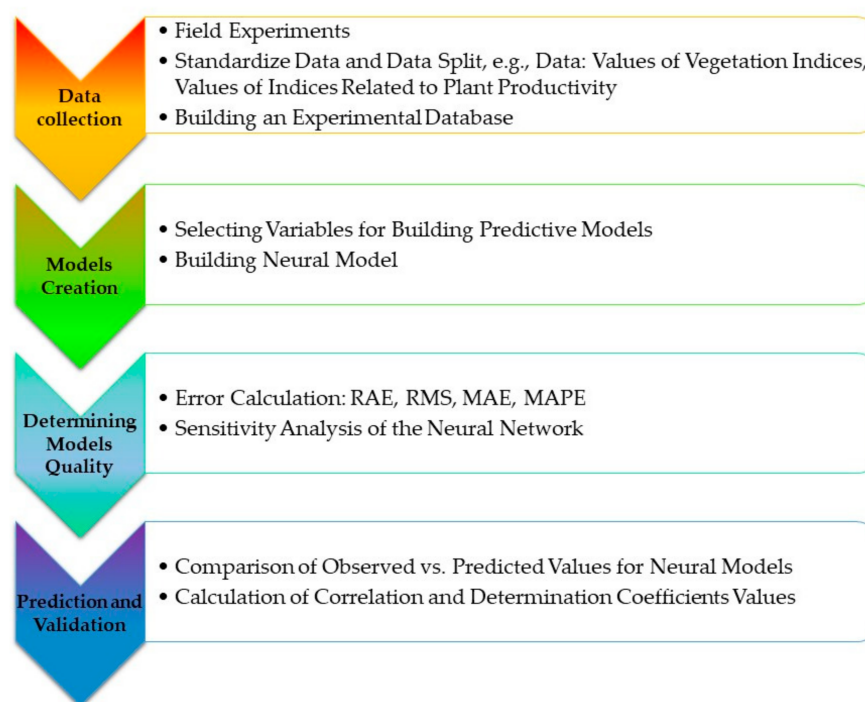
Figure 1. Flowchart showing the steps of working with predictive models.

Over the years, ANNs have been successfully used to forecast the yields of agricultural plants [63–68]. Niedbała [69] experimented with ANN using a multilayer perceptron (MLP) topology to forecast the yield of winter rapeseed. The production fields were situated in Poland, in the southern part of the Opolskie voivodeship. The author obtained a model that allowed forecasting the yield on June 30, having the lowest value of the mean absolute percentage error of 9.43%.The neural model research studies by Ayoubi and Sahrawat [70] explained the variability in biomass and the yield of barley grain at the level of 93% and 89%, respectively. According to Piekutowska et al. [71], in addition to the correct selection of a machine learning algorithm, an important element in creating forecasting models is choosing an appropriate number of predictors that actually shape the yield size. The choice of the independent variables constituting a predictive model is therefore a kind of compromise created by the model developer. First, data should be selected that are available throughout the entire forecast period. Secondly, variables that actually shape the quantity and quality of achieved crops should be considered. Therefore, the author of the predictive model must know the object of one's research perfectly.

The factors influencing crop yield can be divided into primary and secondary. Primary factors include environmental indicators such as temperature, precipitation, insolation, soil pH, soil moisture, nutrients abundance in soil, and agronomic factors such as sowing period [4,16,66,72]. The above-mentioned variables characterize the field research environment or result from local climatic conditions. Secondary factors are those that require additional measurements with the use of specialized devices and sensors in order to be known. These include: the main vegetation indices, for example, normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI); plant growth analysis indices such as leaf area index (LAI), gross primary production (GPP), and evapotranspiration (ET); and indicators showing the relationship between photosynthetic production and biomass growth such as fraction of photosynthetically active radiation (FPAR) [9,11,73]. The following variables enable the determination of the state of the crop at the time of the measurements. Such information is crucial as it indicates the condition of the vegetation, which is sensitive to both abiotic and biotic factors [74].

2. Primary Factors

Agriculture is one of many sectors that is vulnerable to ongoing climate change [75,76]. Changes in precipitation and temperature fluctuations during the growing season may contribute to significant yield losses [77]. As such, environmental factors are commonly used variables in predicting crop yields [4,6,17,72].

The independent variable most often used in ANN models is temperature [78]: minimum temperature (TMN), maximum temperature (TMX), and average temperature. Research on forecasting maize productivity performed in the north-eastern part of South Africa showed the different effects of temperature on crop production depending on the location. TMX significantly influenced the yield of maize in two out of four investigated provinces, and in one province TMN, was largely responsible for shaping maize productivity. Considering the maximum and minimum temperature as independent variables allowed the researchers to create models that obtained from 80% to 95% of the confidence range of the forecasts [6]. Niedbała et al. [51], however, using artificial neural networks, created three models forecasting the yield of rapeseed. In the model forecasting the yield at the end of May, one of the important factors influencing the rapeseed's proper growth and development was the average air temperature in the period from 1 to 31 May. Jiang et al. [79], when forecasting the yield of maize for ten states in the United States, identified the ten best input variables, including minimum, maximum, and average temperature.

Considering temperature in models predicting crop yield is justified, as this factor highly influences the growth and development of crops [80]. Temperature distribution during the growing season has the strongest impact on plant productivity. However, if the plant is properly supplied with water, this impact is reduced [81].

The temperature distribution during the dormancy of winter plants and the value of this parameter at various stages of their vegetation crucially affect yield. Low temperatures in winter are a stress factor for plants. In the case of winter grains, such as wheat, frost damage reduces the number of ears or seeds and, in extreme cases, may cause wheat seedlings to die [82]. Additionally, the exposure of plants to low temperature reduces their height, number of leaves, and the length of internodes [83]. The occurrence of higher-than-average temperatures in the winter season may result in a decrease in the level of plant frost resistance. This phenomenon has a particularly negative impact when severe frosts reoccur because such plants are then less immune to its effects [81,84]. All these abnormalities in plant development not only reduce yield, but also decrease yield quality.

Extreme high temperatures are detrimental to the proper development of plants. Asseng et al. [85] noted the average temperature fluctuations during the wheat growing season. Fluctuations of 2 °C in major regions in Australia resulted in reductions in wheat grain yield of up to 50%. Much of this is caused due to the ageing of the leaves as a result of the high temperatures. Temperature increases in critical phases of grain growth, such as the grain filling stage, can shorten this process. The dry mass of the grain is positively correlated with time of grain maturation. The occurrence of high temperatures during this period negatively impacts the yield by inhibiting the transport of photosynthetic products to the grain [86–88].

An essential environmental factor determining the proper growth and development of plants is water and its availability in soil. Although plants have developed mechanisms at the molecular level that allow for the reduction in resource consumption and adjust their growth to unfavorable environmental conditions, water stress and water deficits still pose a serious threat to agricultural crops [89]. In areas where the sum of precipitation and its distribution do not exceed the limit norms, plants produce a higher yield compared with plants grown in areas with a water deficiency [6,90]. A similar case occurs with excess water. A significant increase in rainfall intensity may deteriorate soil fertility due to blurring of soil colloids, negatively affecting plant production. The sum of rainfall exceeding the capacity of the soil to retain water leads to leaching of nitrogen beyond the root area of plants, thus disrupting its proper development [91]. However, for agriculture, information on the distribution of rainfall from the start of the growing season throughout flowering to

full maturity is important. The water demand of many plant species in the period from sowing to harvesting is often much higher than the sum of precipitation; therefore, both rainfall and its distribution determine the level of plant productivity [92,93]. Given the above, the distribution of precipitation should be considered while conducting research on forecasting models.

Niedbała [69] predicted rapeseed yield from fields located in Poland, in the central and south-western part of the Greater Poland voivodeship. The author, using a multilayer perceptron (MLP) model, showed that the independent variable (the sum of precipitation in the period from 1 September to 31 December) obtained the highest rank for the two models, showing that the indicator had the stronger impact on the yield. Rape is a plant that is sensitive to water stress during key phases such as flowering and seeding. During the flowering period, rapeseed reacts to water stress with a significant delay in reaching maturity. However, subjecting plants to water stress during maturing of grains may result in their earlier ripening [91]. In another study, Niedbała and Kozłowski [94], using the same yield forecasting method (MLP), obtained models forecasting the yield of winter wheat with a low forecast error (MAPE). This error ranged from 8.85% to 9.07% for all three models. In this study, one of the independent variables was precipitation occurring in the period from September to June of the following year. In the next case study [95], the average monthly values of daily precipitation were used as explanatory variables for projecting almond yield. The research was conducted in Central Valley, California, USA, and covered the years 2009 to 2018. The applied random forest model accounted for 82.0% of the almond nut yield variation on average and had an RMSE of 480 ± 9 lbs·acre^{-1}. The inclusion of weekly precipitation data as an independent variable in predicting the yield of soybeans and maize grown in Maryland, USA was reported by Kaul et al. [96]. According to these authors, monthly data may be insufficient to obtain an early and reliable forecast of crop yield.

More frequent extreme weather conditions, especially in terms of changes in precipitation and ambient temperatures, may cause disturbances in the physiological processes of plants [97]. Therefore, in agronomic research, it is important to know the Sielianinov value of the hydrothermal coefficient (HTC). Considering this parameter as an independent variable in the work on predictive models is an interesting and non-standard approach. This coefficient considers the sum of precipitation in a given period (month, quarter) and the sum of the temperatures of this period [98,99]. For this reason, it can be used as a predictor replacing the explanatory variable of precipitation during the growing season of plants. The HTC determines the water relations in the environment. An HTC value <1 indicates the occurrence of drought, whereas HTC $\geq$ 1 means sufficient humidity [100]. Reliable results of this coefficient were obtained when the average value of daily temperatures exceeds 10 °C [101].

Strong temporal variability in rainfall and potential evapotranspiration at the intra-annual scale significantly affects plant growth and development [102]. Evapotranspiration is the result of synergistic interactions between climate, soil, and vegetation [103]. It is strongly influenced not only by the plant type and species composition at a site, but also by the overall economy of available water and energy. ET is used to infer soil moisture [104–107], which can be a valuable input to crop yield estimation models.

An analysis of the available literature on the application of machine learning performed by van Klompenburg et al. [78] showed that the group of traits most often used in yield forecasting was soil information, which consisted of variables such as soil type, pH value, cation-exchange capacity, and soil maps, which provide information about soil nutrients, soil type, and location.

A key factor affecting agricultural crop performance is soil, which is the main source of water, as well as micro- and macro-elements [108]. Therefore, the knowledge of the physico-chemical properties of soil and including them as independent variables in the prediction of yield may improve the accuracy of forecasting. The relationship between the apparent electrical conductivity of the soil (EC$_a$) and topographic measurements (slope,

curvature, and aspect) and the yield of arable crops (maize, soybean, and sorghum) was investigated. The researchers used a feed-forward network with a maximum of ten neurons in the input layer, ten neurons in the hidden layer, and one neuron in the output layer. They proved that EC_a explains the yield variability better than topographic variables, and the highest mean of goodness-of-fit of the obtained model was an R^2 of 0.40 and 0.39, respectively, for the state of Missouri, USA [109]. Abrougui et al. [4] estimated the yield of potatoes grown in an organic system, where the input variables were soil parameters: soil resistance, water and organic carbon content in soil, as well as the microbiological condition of the soil (the numbers of mesophilic and thermophilic bacteria and fungi were determined). The application of a modular feed-forward network with a topology of two hidden layers with seven neurons produced a model with an MSE value of 0.01.

The cation-exchange capacity (CEC) is an important soil feature that describes the availability of nutrients to plants. It indicates the ability of the soil to retain cationic nutrients such as calcium, magnesium, potassium, and ammonium [110]. Soils with a high CEC are characterized by a higher level of organic matter and/or clay content [110]. Soils with a low CEC are usually sandy and/or poor in organic matter [111,112]. Miao et al. [37], in forecasting the yield of maize in the eastern part of Illinois, USA, using an ANN, noticed that CEC is one of the most important soil factors in the analyzed fields, in which MLP and radial basis function (RBF) networks were used. In turn, the maize yield predictive model constructed by Crane-Droesch [113], which contained over 20 soil indicators including CEC, showed that the importance of these variables was low in shaping the yield. The research was performed for nine states (Illinois, Indiana, Iowa, Kentucky, Michigan, Minnesota, Missouri, Ohio, and Wisconsin) in the USA. The importance of independent variables in the predictive processes may differ from each other due to the differences in field location. Each field is characterized by different physico-chemical properties of soil, including the abundance of nutrients. Forecasting the yield of the same plant species, but grown in a different field, may result in obtaining different weights for the same independent variables in the process of training the neural network as other key variables affect the yield. All these aspects vary the crop yield result depending on time and space, as evidenced by the research conducted by Adisa et al. [114], who analyzed the alternations in agroclimatic parameters affecting maize productivity in the north-eastern part of South Africa.

In the research devoted to plant yield forecasting with the use of artificial neural networks, the intensity of solar radiation and wind force have also been used as some of the explanatory variables. The dependence of climatic factors, including wind speed and hours of sunshine, on the yield of rice seeds cultivated in Sri Lanka was analyzed. Three ANN algorithms were used: the Levenberg–Marquardt algorithm, the Bayesian regularization algorithm, and the scaled conjugated gradient algorithm. The research found that all analyzed models produced high-accuracy predictions (the MSE value ranged from 0.01 to 0.39 t·ha^{-1}). However, the Levenberg–Marquardt and scaled conjugated gradient algorithms required fewer epochs and a shorter computation time [115]. Gonzalez-Sanchez et al. [18], apart from the sum of precipitation, minimum and maximum temperature, relative humidity, and field location, also used solar radiation (in MJ/m^2) as an explanatory variable in the prediction of agricultural crops grown in Sinaloa (Western Mexico). The MLP neural network forecasting the yield of snap bean considering the above-mentioned predictors obtained an RMAE ranging from 1.72% to 6.41%. However, the forecast model of the size of maize and tomato was characterized by an error of 8.46% and 24.27%, respectively. The highest error value was obtained for the potato yield forecast model (RMAE was 26.29% on average). In other studies on rice yield forecasting, apart from climatic factors (temperature, precipitation, evaporation, solar radiation, sunshine duration, wind speed, pressure at the station, etc.), the biological features of plants were also included (effective panicle number, filled grains per panicle, and growth period) and agronomic factors (seed set rate). Two models were used in the following research: feed-forward backpropagation neural network (FFBN) and partial least squares regression (PLSR). The area of analysis was fields located in eastern China. The acquired results showed that after

incorporating all the predictive variables, the FFBN model was more accurate. The RMSE values for the training and test sets were 0.41 and 0.44 t·ha^{-1}, respectively. The PLSR model showed an error of 0.56 and 0.55 t·ha^{-1} for the training and test sets, respectively [116].

The agrotechnical treatments performed in the previous year and/or in the year of forecasting are also important input variables that allow the yield to be estimated with satisfactory accuracy. The sowing date is one of the key agrotechnical treatments that is a crucial yield factor. For example, a delay in sowing of maize in the northern part of New Zealand may result in a yield reduction of 10% to 24% depending on the grown cultivar [83]. Using the planting date as one of the inputs enabled Zhang et al. [117] to develop a feed-forward neural network that predicted the phenological development of soybean. The mean prediction error for vegetative development was 3.6 days, and for generative growth, it was 4.4 days. The results of the sensitivity analysis of the neural network showed that the sowing date is a core independent variable in forecasting the yield of rapeseed. Moreover, including the fertilization doses of nitrogen, potassium, phosphorus, magnesium, molybdenum, zinc, sulfur, and copper enabled the construction of prognostic models whose MAE ranged from 0.52 to 0.55 t·ha^{-1}, which correspond to MAPE values of 6.63–6.92% [118].

Fertilization is one of the agrotechnical procedures used to provide plants with digestible forms of nutrients. The natural concentration of the above-mentioned compounds in the growing environment is usually insufficient. The application of mineral and organic fertilizers, in proper doses, affects the correct growth and development of plants. Too-low fertilization leads to deficiencies in micro- and macro-elements in the plant, thus disrupting its physiological processes. It also contributes to soil depletion from absorbable forms of nutrients. In turn, large doses of fertilizers, exceeding plant requirements, also disturb the ionic balance in soil [119,120]. In sustainable agriculture, it is particularly important that the doses of fertilizers are adjusted to the agrochemical properties of the soil and the nutritional requirements of the cultivated plants [121,122]. Niedbała et al. [69] noted that mineral fertilization is one of the important agrotechnical features in forecasting the yield of wheat. The analysis covered crops located in Poland in the central and south-western part of Greater Poland. Although the weather conditions were ranked first in the obtained models, by including fertilization, models with MAPEs ranging from 6.63% to 6.92% were obtained. The researchers used an MLP network with the structure 23:38-16-8-1:1.

The number of primary factors used as independent variables in plant yield prediction, and presented in this study, validates the high complexity of the task. The most common predictors in ANN modeling were outlined in the section above. However, the correctness of the prediction model depends not only on the quality of the data but also on the representativeness of the model. Data with outliers, incomplete sets, or erroneous significantly limit the forecasting model capabilities [123,124].

3. Indices Related to Plant Productivity

Plant productivity indices are directly related to the concept of remote sensing, which is defined as "the study of obtaining information about an object by analyzing data from a device that is not in contact with the object" [125]. Such data can be obtained from devices such as sensors, digital cameras, and video recorders, which are placed on various platforms (airplanes, satellites, unmanned aerial vehicles, or handheld radiometers) and obtain data in various forms, including the distribution of acoustic waves or the type of electromagnetic energy [126]. In remote sensing, land cover monitoring has become one of the most active areas of research [127,128]. The development of forecasting models using remote sensing data is a solution with high use potential. These data provide quantitative and up-to-date information on the development of crops over a large area in a cost-effective manner. Moreover, the advantage of such measurements is their non-invasive nature, which means that they can be obtained without the need to destroy plant tissue [9,129–132]. Plant productivity indices may be divided into the following [133–135]:

1.　Growth analysis indices, i.e., leaf area index (LAI), leaf area ratio (LAR), relative growth rate (RGR), leaf area duration (LAD), unit leaf rate (ULR), and weighted difference vegetation index (WDVI);
2.　Indices that show the relationship between photosynthetic production and biomass growth, i.e., photosynthetically active radiation (PAR), fraction of photosynthetically active radiation (FPAR), radiation use efficiency (RUE), and gross primary production (GPP);
3.　Vegetation indicators, i.e., normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), enhanced vegetation index 2 (EVI 2), and ratio vegetation index (RVI);
4.　Indices determining the content of chlorophyll, i.e., MERIS terrestrial chlorophyll index (MTCI), triangular vegetation index (TVI), modified triangular vegetation index 1 (MTVI 1), modified triangular vegetation index 2 (MTVI 2), chlorophyll absorption reflectance (CAR), canopy chlorophyll content index (CCCI), etc.

The NDVI, EVI, and LAI indices are some of the most commonly used plant productivity indices. Their calculation methods are presented in the following formulas:

$$\text{NDVI} = \frac{NIR - RED}{NIR + RED} \tag{3}$$

where **NIR** and **RED** are the reflectance in the near infrared (**NIR**) and red bands, respectively [136];

$$\text{EVI} = G \cdot \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + C_1 \cdot \rho_{red} - \rho_{blue} + L} \tag{4}$$

where **G** is a scaling factor; ρx is the atmospherically corrected surface reflectance, differential **NIR**, and red radiative transfer through a canopy; C1 and C2 are the coefficients of the aerosol resistance term; and L is the canopy background adjustment for correcting nonlinearity [137];

$$\text{LAI} = \frac{A_c}{P} \tag{5}$$

where A_c is the total area of the leaves of the whole canopy and **P** is the ground area occupied by the plants [81].

The NDVI determines plant vitality and photosynthetic activity and is calculated from the reflection of light in the near infrared and red-light bands. The EVI, however, is calculated by turning on the reflection of the blue band to solve the saturation problem of the NDVI. The LAI is defined as the total leaf area per unit ground surface area and is used as an approximation of the leaf biomass. LAI is also used to model the evapotranspiration of crops. Relatively commonly in estimating the yield of crops, signals are used to determine the content of chlorophyll in plants, such as CCCI [5,126,138], which is calculated as

$$\text{CCCI} = \frac{NDRE - NDRE_{min}}{NDRE_{max} - NDRE_{min}} \tag{6}$$

where **NDRE** is the normalized difference red edge; and **NDRE$_{min}$** and **NDRE$_{max}$** are the minimum and maximum values of this index, respectively [139].

Over the past few years, there has been growing interest among researchers in the practical application of indices related to plant productivity [140–144]. For instance, to support the assessment of the condition of potato crops, spectral data obtained from the Sentinel-2 satellite were used [73]. The research data covered farmland in the south of the Netherlands, which, after being obtained from the satellite, were compared with the results obtained from measurements with a manual radiometer. The analysis indicated that satellite data can be successfully used to determine parameters such as LAI, CCCI, and WDVI. A different study used data from the Sentinel-2 satellite to predict the yield of commercial potato tubers grown in Segovia, Spain. Eight different machine learning algorithms were applied in the research: Support vector machine radial, random forest,

k-nearest neighbors, multivariate adaptive regression splines, model averaged neural network, quantile regression with LASSO penalty, and linear regression with backward selection. The following parameters were applied in the study: the anthocyanin reflectance index, which estimates the content of anthocyanins; the carotenoid reflectance index, which assesses the content of carotenoids; the inverted red-edge chlorophyll index, which determines the content of chlorophyll in the canopy; and the leaf chlorophyll content, which determines the concentration of chlorophyll in a leaf area unit. Additionally, the researchers used the NDVI; the plant senescence reflectance index (PSRI), reflecting the senescence of leaves; and the weighted difference vegetation index, which is a parameter similar to LAI. All measurements were recorded between the tuberization period and the ageing of the plants. Despite the lack of results related to weather and soil properties, researchers managed to create a model (support vector machine radial) that was able to forecast tuber yield several weeks before harvest. The mentioned model had an MAE of 8.64% and RMSE of 11.70% [132]. Leaf area index (LAI), chlorophyll content in leaves, and different nitrogen fertilization levels were used as input data to forecast winter wheat grain yield [145]. The wheat plantation was located in Belm, in the northwest of Germany, and was grown in the second half of October 2016, whereas LAI and SPAD measurements (later used to calculate the content of chlorophyll ($\mu g \cdot cm^{-2}$)) were recorded on 20 June 2017. The MLR model was used in the research to estimate the yield size. The RMSE of the model was on average 4183 $dt \cdot ha^{-1}$. In another case study, Rahman et al. [146] used, inter alia, the NDRE index to predict the yield of mango fruit. Plant productivity indices were measured in the early phase of fruition of trees. The study used a feed-forward model with backpropagation, and the covered orchards were located in Northern Australia. The combination of plant vegetation indices, including NDRE and tree crown area, allowed the authors to obtain an ANN model an R^2 of 0.7. The RMSE for the total fruit weight was 13.83 $kg \cdot tree^{-1}$.

Abbas et al. [7] forecasted the yield of potato tubers, in Canada on Prince Edward Island and in New Brunswick. In the following studies, independent soil parameters were used as variables: electrical conductivity, organic matter content, volumetric moisture, soil pH, and cation-exchange capacity. The NDVI was also used in the study, which was measured at the end of July (60 days after planting potatoes), in mid-August (80 days after planting potatoes), and at the end of August, in each study year. By including soil data and the plant vegetation index, the authors obtained models with RMSEs ranging from 4.62 to 6.60 $t \cdot ha^{-1}$. The NDVI was also used as a predictor in forecasting sugar cane yield in Brazil [142]. The ANN-based predictive model forecasted the yield three months before harvest, and the RRMSE did not exceed 8%. Kross et al. [12] studied the relationship between the topography of the area, NDVI, NDVIre, and simple ratio (SR) indices and the yield of maize and soybeans. These crops were grown in Eastern Ontario, Canada, and remote sensing data were acquired from June to August of each study year. The research proved that the MLP network developed by the authors was more effective in predicting the yield of maize than that of soybean. The RMAE for the forecasting of maize for the two tested years (2011–2012) did not exceed 15%. In contrast, the RMAE for soybeans was less than 20%. Serele et al. [147] used a back-propagation error ANN to predict maize seed yield. The network was trained using the conjugate gradient method algorithm in farmland located near Ottawa, Canada. In the above study, the authors used independent topographic features (slope and aspect), vegetation indices (including NDVI, WDVI, and SAVI) and textural indices (including homogeneity, contrast, and entropy) as variables. Various combinations of these variables were analyzed in the development of the predictive model. The research showed that the model containing all explanatory variables was more accurate than the other models. The RMSEs for the training and validation set were 0.36 and 0.42 $t \cdot ha^{-1}$, respectively. Panda et al. [141] also tested predictive models that included different vegetation indices as independent variables. In the work on forecasting the yield of maize cultivated in North Dakota (USA), four indicators were considered: NDVI, global vegetation index (GVI), perpendicular

vegetation index (PVI), and soil-adjusted vegetation index (SAVI). Back-propagation neural network models were developed, producing a total of 16 models (four vegetation index × four years of research, including data from interconnected years). The model that used PVI as a predictor exhibited the best forecast accuracy compared with the other models. The average accuracy of the estimation of the size of maize yields was 83.5%, 93.0%, and 96.0% for 1998, 1999, and 2001, respectively. According to the authors, the high accuracy of the model was due to the PVI being better at reducing noise caused by bare soil information present in spectral images. Feng et al. [148] used 40 indices related to plant productivity, including MTCI, NDVI, and EVI, to predict alfalfa (lucerne) yield, located in the state of Wisconsin (north-central United States). The crop was sown in May 2018 and 2019, whereas the measurements were recorded on 25 July and 19 August 2019. The most accurate model obtained by the researchers was characterized by a R^2 of 0.87. As demonstrated above, progress in remote sensing techniques has allowed for the application of multispectral images as an effective tool for predicting plant productivity [149].

4. Restrictions on Selected Input Variables

Each independent variable used in predictive models, such as measurements of air temperature, relative air humidity, sum of precipitation, or soil physicochemical properties, may be limited due to human error or failure of the measuring devices. Errors in remote sensing data are often caused by weather and ground conditions (e.g., overcast and snow cover) and/or sensor problems (e.g., sensor drift and changes in sensor view angle). These errors may result in irregularly low data values [150]. The accuracy of measurements is important for ensuring reliable results [151]. Accuracy is in line with the term uncertainty, which covers a wider range of doubts or inconsistencies in the obtained data [152]. One source of uncertainty may be the subjectivity of data collection [153]. Limiting the occurrence of these conditions has important impacts on the accuracy of the model. Detecting data inaccuracies, also referred to as fraud detection [154], can contribute to increasing yield prediction accuracy. The identification of such data allows them to be ignored in modeling, which limits the amount of information that the neural network has to process. This may, in turn, shorten the time needed to generate a forecast and reduce the data that lower the model's accuracy. Nevavuori et al. [3] used an unmanned aerial vehicle to obtain NDVI and RGB data from fields located in the southwestern part of Philadelphia, USA, which were used to evaluate which of these indices produces better results in wheat and barley yield prediction using convolutional neural networks (CNNs). The results showed that using the RGB image predictor early in the growth phase (<25% of the total thermal time) produced a better functioning model compared with the NDVI model. The RGB image model was characterized by a mean absolute error of 0.48 t·ha^{-1} (MAPE was approx. 8.8%). The authors suggested that using deep network learning to forecast yield from drone photos may be useful as long as these images are taken relatively early in the season [3]. The NDVI does not have a feedback loop (open loop structure), which makes it susceptible to numerous errors and uncertainty given changing weather conditions and the background of the canopy. However, this problem was solved by the development of a modified NDVI indicator in the form of the EVI. This parameter is characterized by a higher resistance to canopy background noise and weather conditions [155,156]. The higher efficiency of EVI in characterizing plant productivity makes it a more effective indicator in crop forecasting in comparison with NDVI. This is confirmed by Bolton and Friedl [157], who estimated the yield of maize and soybeans in central United States. Although EVI is able to reduce the noise associated with the influence of the atmosphere and the background of the canopy, it does not include topographic effects, which are defined as "the change in radiation accompanying the change of orientation from a horizontal into an inclined surface in response to a change in the position of the light source" [158]. The following effect is another important environmental factor influencing the formation of noise in the calibration of vegetation indices, which is of particular importance in mountainous areas [155]. Interestingly, Johnson et al. [9] demonstrated that in forecasting the yield of barley, wheat,

and rapeseed, the NDVI was more accurate than the EVI. The NDVI appeared to be a more efficient vegetation index in forecasting crops in the Canadian prairies. The contradictions in these results are caused by the variability in plant biomass. In various phenological phases, the vegetation indices may have different values due to differences in plant growth, which are determined by changing weather conditions [159]. This was confirmed by the results by Son et al. [156], who obtained different correlation coefficients for NDVI and EVI indices in rice cultivation. These parameters, as shown by the research, were determined by climatic conditions. Furthermore, according to Zhang and Zhang [160], NDVI may be more effective in some regions, whereas EVI may be more so in others. The efficiency of various vegetation indices may also result from the location of the analyzed areas. For instance, the normalized difference water index (NDWI) is more sensitive to irrigation in semi-arid areas with low agricultural density compared with NDVI and EVI2 [157].

The spatial resolution of the image is an important factor influencing the quality of the acquired satellite data. Ensuring the appropriate distance of the satellite from the Earth's surface during measurements is essential for obtaining high-quality data [161]. The Advanced Very High Resolution Radiometer (AVHRR) is an NOAA platform device [162] with a spatial resolution of 1 km [163]. The spatial resolution of the Moderate Resolution Imaging Spectroradiometer (MODIS) on the Terra and Aqua satellites is much higher (up to 250 m) compared with the AVHRR [164]. Both radiometers are used to obtain information on plant vegetation indices. However, due to the higher spatial resolution, more effective results are acquired with MODIS [165]. Additionally, the two main limitations of AVHRR are the overlapping of the near-infrared channel with the water vapor absorption region of the atmosphere, leading to noise in the remote sensing signal and the relatively fast saturation of the red channel and, thus, of NDVI [166]. The above limitations affect the accuracy of neural predictive models. The quality of the obtained ANN model was assessed by Li et al. [167], who found that MODIS-NDVI is more accurate than AVHRR-NDVI in predicting the yield of soybeans and maize grown in the corn belt area, located in the North American Midwest and covers nine states in USA. Similar results were also obtained by Mkhabela et al. [168] in the forecast of barley, rape, pea, and wheat crops, which were performed in the Canadian Prairies. These studies differed in the method of forecast used by the authors: linear regression methods were used for predicting yields, and not ANN. According to Chen et al. [169], remote sensing data with a higher spatial resolution are necessary for the effective detection and monitoring of changes in the entire landscape. Their improvement would also produce measurable benefits in precision agriculture, including the prediction of plant productivity.

Limitations in the use of remote sensing data in forecasting agricultural crops result mainly from the large scale of the study area. Weather data are usually collected on the micro-environment scale [170], whereas remote sensing data, if they are not generated by unmanned aerial vehicles but by satellites, are obtained from macro-space. The aforementioned increase the disturbances in data collection [12]; as a result, the large area covered by remote sensing analysis causes each pixel of the plant productivity index to contain information about all crops in that area. Plant productivity indices mainly concern the dominant crops, whereas the less frequently cultivated plants are ignored [9].

5. Current Trends in Creating Forecasting Models

Statistical models are simple to use and less demanding in terms of input variables. However, they are highly limited with respect to information they provide beyond the range of values for which the model is parameterized [126]. In addition, these models are often criticized for failing to provide a scientific understanding of the processes studied [13]. Therefore, in the near future, the interest of scientists, farmers, and decision makers will focus on machine learning, including artificial neural networks.

Given the current knowledge and technology, one of the problems is the selection of an appropriate learning and forecasting method adapted to a specific problem and data set. According to research by Zhang et al. [171], the selection of the correct method of training

neural networks and the method of forecasting the grain yield of rice has crucial effects on the accuracy of prediction. The study considered fields located in the northern, central and southern parts of Burkina Faso. Three different forecasting methods were used: ANN, conventional multiple regression, and boosted regression trees. Furthermore, four different neural network learning algorithms were used: multilayer perceptron, probabilistic neural network, and generalized feed-forward and linear regression. Among the forecasting methods, the multiple regression model attained the highest MAE (0.34 t·ha^{-1}). In turn, among the neural network learning techniques, the ANN linear regression model was characterized by the largest MAE, which was the same as for the multiple regression model, i.e., 0.34 t·ha^{-1}. The probabilistic ANN model was characterized by the lowest error rate (MAE = 0.12 t·ha^{-1}). Khaki and Wang [172] obtained similar results using a less popular approach of forecasting plant productivity, as they covered 2247 fields located in the United States and Canada. For the prediction of maize yield, four different models were applied: deep neural network (DNN), least absolute shrinkage and selection operator (LASSO), shallow neural network (having a single hidden layer with 300 neurons), and regression tree. The achieved cross-validation results indicated that the most accurate model was the DNN model, with an RMSE of 12.79 dt·ha^{-1}. The LASSO model was the least accurate, with an average RMSE of 21.40 dt·ha^{-1}. The further development of artificial neural networks in the forecasting of agricultural crops should be targeted toward determining to what extent this approach can be implemented and developed in precision agriculture [173]. The current trends in yield forecasting focus on remote sensing data that reflect the condition of the crop. Recent scientific works have demonstrated the possibility of estimating crop yield based on hyperspectral data combined with weather data. For example, Kuwata and Shibasaki [174] predicted the United States corn yield using independent weather data and satellite plant productivity indices as variables. The best yield prediction results were achieved using a deep neural network model. Kim et al. [5] also used remote sensing data and weather information to assess the size of maize and soybean crops in the Midwestern United States. The following predictors were used in these studies: NDVI; EVI; LAI; FPAR; GPP; minimum, maximum, and average temperature; and sum of precipitation. The results highlighted that the ANN model had lower prediction accuracy compared with deep neural networks, whose prediction error was on average 7.6% and 7.8% (for maize and soybean, respectively). The correlation coefficient (r) for the created model was 0.95 for maize and 0.90 for soy. Pantazi et al. [10] examined the relationship between the NDVI obtained from the UK-DMC-2 platform and soil parameters (pH, organic matter content, soil moisture, content of calcium, magnesium, phosphorus in the soil, CEC, and others) and the yield of winter wheat. The three models based on self-organizing maps forecasted yield with an accuracy of 91.30% to 92.15%. Hyperspectral and weather data are now widely used in predictive models because these data have become easily available to professional users of precision farming. In addition, the data can be downloaded and analyzed at any stage of the growing season, proving their utility in pre-harvest yield forecasts.

Future research on the application of various types of environmental data in yield forecasting may focus not only on the hybrid use of remote sensing and weather data, but also on the search for new, more reliable indicators of plant productivity that will significantly improve the accuracy of prediction. Cai et al. [74] proposed the use of solar-induced chlorophyll fluorescence (SIF) from a specific NIR band as one of the predictors in the prediction of wheat yield in Australia. The achieved results showed that the combination of climate and satellite data produces higher-quality prediction compared with using only weather data or only satellite data. However, the application of EVI + SIF + climate as independent variables resulted in the same model efficiency as using EVI + climate as the input. Failure of the SIF signal to provide unique information is explained by these data being sparse in relation to time (one-month time period). In addition, SIF data have high spatial resolution and therefore cannot capture small traits in space [135,175]. Possibly, the application of SIF with better spatial resolution, which can be acquired by NASA's OCO-2 satellite, would affect the accuracy of the yield forecast.

Research studies directed towards dynamic agricultural modeling should be continued and developed to allow for an in-depth assessment of the efficiency of neural prediction models from a wider perspective than previously: the local environment, crop productivity, and economic effects [16].

6. Summary

The application of nonlinear methods of yield forecasting is necessary due to the complexity of the agricultural system. Numerous environmental factors responsible for shaping the efficiency of plant yield, and their nonlinear nature requires departing from traditional statistical modeling methods in favor of more precise prediction methods. Models based on artificial neural networks are a suitable alternative; though ANNs are widely used in yield forecasting, their practical application still faces some difficulties. The most important are the following: selection of an appropriate number of hidden network layers, speed of model training, and the application of a sufficient amount of data in the form of independent variables. However, ANNs currently play a key role in precision agriculture. The results of the analyses obtained through the application of this tool contribute to the increase in the profitability of farms.

ANN data are a source of necessary and reliable information in agricultural production management. Information about the yield can be obtained even a few months before harvest, which is extremely valuable for adopting an appropriate strategy in the import and export of agricultural products. In addition, prior knowledge of the yield of crops allows rationalizing the production means, which is in line with the idea of sustainable development.

The number of factors influencing crop yield and the parameters describing the condition of the canopy complicate the selection of those appropriate for a given crop. Weather data and information on agricultural technology are two of the most crucial predictors influencing the accuracy of crop forecast.

We think that by identifying the most frequently used independent variables in yield prediction, this article will be helpful for many researchers in future studies.

Author Contributions: Conceptualization, P.H., M.P. and G.N.; methodology, P.H. and M.P.; 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.; funding acquisition, M.P. and G.N. All authors have read and agreed to the published version of the manuscript.

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

Abbreviations

ANN—artificial neural networks; AVHRR—advanced very high resolution radiometer; CAR—chlorophyll absorption reflectance; CCCI—canopy chlorophyll content index; PSRI—plant senescence reflectance index; CEC—cation-exchange capacity; CNNs—convolutional neural networks; DNN—deep neural networks; EC_a—electrical conductivity of the soil; ET—evapotranspiration; EVI—enhanced vegetation index; EVI 2—enhanced vegetation index 2; FFBN—feed-forward backpropagation neural network; FPAR—fraction of photosynthetically active radiation; GPP—gross primary production; GVI—global vegetation index; HTC—hydrothermal coefficient; LAD—leaf area duration; LAI—leaf area index; LAR—leaf area ratio; RGR—relative growth rate; MAPE—mean absolute percentage error; MLP—multilayer perceptron; MLR—multiple linear regression; MODIS—moderate resolution imaging spectroradiometer; MSE—mean squared error; MTCI—MERIS terrestrial chlorophyll index; MTVI 1—modified triangular vegetation index 1; MTVI 2—modified triangular vegetation index 2; NDRE—normalized difference red Edge; NDVI—normalized difference vegetation index; NIR—near infrared; PAR—photosynthetically active radiation; PLSR—partial least squares regression; PVI—perpendicular vegetation index; RAE—relative absolute error; RBF—radial basis function; RMSE—root mean square error; MAE—mean absolute error; RRSE—root relative squared

error; RUE—radiation use efficiency; RVI—ratio vegetation index; SAVI—soil-adjusted vegetation index; SIF—solar-induced chlorophyll fluorescence; SPAD—soil plant analysis development; TMN—minimum temperature; TMX—maximum temperature; TVI—triangular vegetation index; ULR—unit leaf rate; WDVI—weighted difference vegetation index.

References

1. Rose, D.C.; Sutherland, W.J.; Barnes, A.P.; Borthwick, F.; Ffoulkes, C.; Hall, C.; Moorby, J.M.; Nicholas-Davies, P.; Twining, S.; Dicks, L.V. Integrated farm management for sustainable agriculture: Lessons for knowledge exchange and policy. *Land Use Policy* **2019**, *81*, 834–842. [CrossRef]
2. 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]
3. Nevavuori, P.; Narra, N.; Lipping, T. Crop yield prediction with deep convolutional neural networks. *Comput. Electron. Agric.* **2019**, *163*, 104859. [CrossRef]
4. Abrougui, K.; Gabsi, K.; Mercatoris, B.; Khemis, C.; Amami, R.; Chehaibi, S. Prediction of organic potato yield using tillage systems and soil properties by artificial neural network (ANN) and multiple linear regressions (MLR). *Soil Tillage Res.* **2019**, *190*, 202–208. [CrossRef]
5. Kim, N.; Ha, K.-J.; Park, N.-W.; Cho, J.; Hong, S.; Lee, Y.-W. A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006–2015. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 240. [CrossRef]
6. Adisa, O.; Botai, J.; Adeola, A.; Hassen, A.; Botai, C.; Darkey, D.; Tesfamariam, E. Application of Artificial Neural Network for Predicting Maize Production in South Africa. *Sustainability* **2019**, *11*, 1145. [CrossRef]
7. Abbas, F.; Afzaal, H.; Farooque, A.A.; Tang, S. Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms. *Agronomy* **2020**, *10*, 1046. [CrossRef]
8. Strapatsa, A.V.; Nanos, G.D.; Tsatsarelis, C.A. Energy flow for integrated apple production in Greece. *Agric. Ecosyst. Environ.* **2006**, *116*, 176–180. [CrossRef]
9. 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]
10. Pantazi, X.E.; Moshou, D.; Alexandridis, T.; Whetton, R.L.; Mouazen, A.M. Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* **2016**, *121*, 57–65. [CrossRef]
11. Kim, N.; Lee, Y.-W. Machine Learning Approaches to Corn Yield Estimation Using Satellite Images and Climate Data: A Case of Iowa State. *J. Korean Soc. Surv. Geod. Photogramm. Cartogr.* **2016**, *34*, 383–390. [CrossRef]
12. Kross, A.; Znoj, E.; Callegari, D.; Kaur, G.; Sunohara, M.; Lapen, D.R.; McNairn, H. Using artificial neural networks and remotely sensed data to evaluate the relative importance of variables for prediction of within-field corn and soybean yields. *Remote Sens.* **2020**, *12*, 2230. [CrossRef]
13. Qian, B.; De Jong, R.; Warren, R.; Chipanshi, A.; Hill, H. Statistical spring wheat yield forecasting for the Canadian prairie provinces. *Agric. For. Meteorol.* **2009**, *149*, 1022–1031. [CrossRef]
14. Guo, W.W.; Xue, H. An incorporative statistic and neural approach for crop yield modelling and forecasting. *Neural Comput. Appl.* **2012**, *21*, 109–117. [CrossRef]
15. Drummond, S.T.; Sudduth, K.A.; Joshi, A.; Birrell, S.J.; Kitchen, N.R. Statistical and neural methods for site–specific yield prediction. *Trans. ASAE* **2003**, *46*, 5. [CrossRef]
16. Khairunniza-Bejo, S.; Mustaffha, S.; Ismail, W.I.W. Application of artificial neural network in predicting crop yield: A review. *J. Food Sci. Eng.* **2014**, *4*, 1–9.
17. Niedbała, G.; Nowakowski, K.; Rudowicz-Nawrocka, J.; Piekutowska, M.; Weres, J.; Tomczak, R.J.; Tyksiński, T.; Pinto, A.Á. Multicriteria prediction and simulation of winter wheat yield using extended qualitative and quantitative data based on artificial neural networks. *Appl. Sci.* **2019**, *9*, 2773. [CrossRef]
18. 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*, 1–10. [CrossRef]
19. Khoshnevisan, B.; Rafiee, S.; Omid, M.; Mousazadeh, H. Development of an intelligent system based on ANFIS for predicting wheat grain yield on the basis of energy inputs. *Inf. Process. Agric.* **2014**, *1*, 14–22. [CrossRef]
20. Khoshnevisan, B.; Rafiee, S.; Omid, M.; Mousazadeh, H. Prediction of potato yield based on energy inputs using multi-layer adaptive neuro-fuzzy inference system. *Measurement* **2014**, *47*, 521–530. [CrossRef]
21. Amid, S.; Mesri Gundoshmian, T. Prediction of output energies for broiler production using linear regression, ANN (MLP, RBF), and ANFIS models. *Environ. Prog. Sustain. Energy* **2017**, *36*, 577–585. [CrossRef]
22. Vivas, E.; Allende-Cid, H.; Salas, R. A Systematic Review of Statistical and Machine Learning Methods for Electrical Power Forecasting with Reported MAPE Score. *Entropy* **2020**, *22*, 1412. [CrossRef]
23. Wang, X.; Huang, J.; Feng, Q.; Yin, D. Winter Wheat Yield Prediction at County Level and Uncertainty Analysis in Main Wheat-Producing Regions of China with Deep Learning Approaches. *Remote Sens.* **2020**, *12*, 1744. [CrossRef]
24. Zhao, Y.; Potgieter, A.B.; Zhang, M.; Wu, B.; Hammer, G.L. Predicting Wheat Yield at the Field Scale by Combining High-Resolution Sentinel-2 Satellite Imagery and Crop Modelling. *Remote Sens.* **2020**, *12*, 1024. [CrossRef]

25. Felipe Maldaner, L.; de Paula Corrêdo, L.; Fernanda Canata, T.; Paulo Molin, J. Predicting the sugarcane yield in real-time by harvester engine parameters and machine learning approaches. *Comput. Electron. Agric.* **2021**, *181*, 105945. [CrossRef]

26. Sharma, L.K.; Singh, T.N. Regression-based models for the prediction of unconfined compressive strength of artificially structured soil. *Eng. Comput.* **2018**, *34*, 175–186. [CrossRef]

27. 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]

28. Kim, S.; Kim, H. A new metric of absolute percentage error for intermittent demand forecasts. *Int. J. Forecast.* **2016**, *32*, 669–679. [CrossRef]

29. Bhojani, S.H.; Bhatt, N. Wheat crop yield prediction using new activation functions in neural network. *Neural Comput. Appl.* **2020**, *32*, 13941–13951. [CrossRef]

30. Singh, R.; Umrao, R.K.; Ahmad, M.; Ansari, M.K.; Sharma, L.K.; Singh, T.N. Prediction of geomechanical parameters using soft computing and multiple regression approach. *Measurement* **2017**, *99*, 108–119. [CrossRef]

31. Chen, J.-F.; Do, Q.; Nguyen, T.; Doan, T. Forecasting Monthly Electricity Demands by Wavelet Neuro-Fuzzy System Optimized by Heuristic Algorithms. *Information* **2018**, *9*, 51. [CrossRef]

32. Gandhi, N.; Petkar, O.; Armstrong, L.J. Rice crop yield prediction using artificial neural networks. In Proceedings of the 2016 IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR), Chennai, India, 15–16 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 105–110.

33. Gandhi, N.; Armstrong, L.J.; Petkar, O.; Tripathy, A.K. Rice crop yield prediction in India using support vector machines. In Proceedings of the 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE), Khon Kaen, Thailand, 13–15 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–5.

34. Schwalbert, R.A.; Amado, T.; Corassa, G.; Pott, L.P.; Prasad, P.V.V.; Ciampitti, I.A. Satellite-based soybean yield forecast: Integrating machine learning and weather data for improving crop yield prediction in southern Brazil. *Agric. For. Meteorol.* **2020**, *284*, 107886. [CrossRef]

35. Mishra, S.; Paygude, P.; Chaudhary, S.; Idate, S. Use of data mining in crop yield prediction. In Proceedings of the 2018 2nd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 19–20 January 2018; pp. 796–802.

36. 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; pp. 44–48.

37. Miao, Y.; Mulla, D.J.; Robert, P.C. Identifying important factors influencing corn yield and grain quality variability using artificial neural networks. *Precis. Agric.* **2006**, *7*, 117–135. [CrossRef]

38. 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]

39. Niazian, M.; Niedbała, G. Machine Learning for Plant Breeding and Biotechnology. *Agriculture* **2020**, *10*, 436. [CrossRef]

40. Liakos, K.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors* **2018**, *18*, 2674. [CrossRef]

41. Dogan, Z.; YILMAZ, M.; BILGILI, A.V. The use of Artificial Neural Networks (ANN) for prediction of time series monthly air temperature and assessment of different neuron numbers on the prediction accuracy. *Fresenius Environ. Bull.* **2015**, *24*, 325–3265.

42. Getahun, M.A.; Shitote, S.M.; Abiero Gariy, Z.C. Artificial neural network based modelling approach for strength prediction of concrete incorporating agricultural and construction wastes. *Constr. Build. Mater.* **2018**, *190*, 517–525. [CrossRef]

43. Khehra, B.S.; Pharwaha, A.P.S. Classification of Clustered Microcalcifications using MLFFBP-ANN and SVM. *Egypt. Inform. J.* **2016**, *17*, 11–20. [CrossRef]

44. Azadeh, A.; Ghaderi, S.F.; Tarverdian, S.; Saberi, M. Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. *Appl. Math. Comput.* **2007**, *186*, 1731–1741. [CrossRef]

45. Karamirad, M.; Omid, M.; Alimardani, R.; Mousazadeh, H.; Heidari, S.N. ANN based simulation and experimental verification of analytical four- and five-parameters models of PV modules. *Simul. Model. Pract. Theory* **2013**, *34*, 86–98. [CrossRef]

46. Liu, M.; Lu, J. Support vector machine—an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river? *Environ. Sci. Pollut. Res.* **2014**, *21*, 11036–11053. [CrossRef] [PubMed]

47. Smuga-Kogut, M.; Kogut, T.; Markiewicz, R.; Słowik, A. Use of Machine Learning Methods for Predicting Amount of Bioethanol Obtained from Lignocellulosic Biomass with the Use of Ionic Liquids for Pretreatment. *Energies* **2021**, *14*, 243. [CrossRef]

48. Tadeusiewicz, R. *Elementarne Wprowadzenie do Techniki Sieci Neuronowych z Przykładowymi Programami*; Akademicka oficyna wydawnicza PLJ: Warszawa, Poland, 1998.

49. Tadeusiewicz, R.; Szaleniec, M. *Leksykon Sieci Neuronowych*; Fundacja na Rzecz Promocji Nauki Polskiej: Wrocław, Poland, 2015.

50. Caselli, M.; Trizio, L.; de Gennaro, G.; Ielpo, P. A Simple Feedforward Neural Network for the PM10 Forecasting: Comparison with a Radial Basis Function Network and a Multivariate Linear Regression Model. *Water. Air. Soil Pollut.* **2009**, *201*, 365–377. [CrossRef]

51. 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]

52. Singh, R.K. Artificial Neural Network Methodology for Modelling and Forecasting Maize Crop Yield. *Agric. Econ. Res. Rev.* **2008**, *21*, 5–10.

53. Ustaoglu, B.; Cigizoglu, H.K.; Karaca, M. Forecast of daily mean, maximum and minimum temperature time series by three artificial neural network methods. *Meteorol. Appl.* **2008**, *15*, 431–445. [CrossRef]

54. Gonzalez-Sanchez, A.; Frausto-Solis, J.; Ojeda-Bustamante, W. Predictive ability of machine learning methods for massive crop yield prediction. *Spanish J. Agric. Res.* **2014**, *12*, 313–328. [CrossRef]

55. Shastry, K.A.; Sanjay, H.A.; Deshmukh, A. A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction. *J. Artif. Intell.* **2015**, *9*, 23–32. [CrossRef]

56. Niazian, M.; Sadat-Noori, S.A.; Abdipour, M. Artificial neural network and multiple regression analysis models to predict essential oil content of ajowan (Carum copticum L.). *J. Appl. Res. Med. Aromat. Plants* **2018**, *9*, 124–131. [CrossRef]

57. Aditya Shastry, K.; Sanjay, H.A. Hybrid prediction strategy to predict agricultural information. *Appl. Soft Comput.* **2021**, *98*. [CrossRef]

58. Zaefizadeh, M.; Jalili, A.; Khayatnezhad, M.; Gholamin, R.; Mokhtari, T. Comparison of multiple linear regressions (MLR) and artificial neural network (ANN) in predicting the yield using its components in the hulless barley. *Adv. Environ. Biol.* **2011**, *5*, 109–113.

59. Niazian, M.; Sadat-Noori, S.A.; Abdipour, M. Modeling the seed yield of Ajowan (*Trachyspermum ammi* L.) using artificial neural network and multiple linear regression models. *Ind. Crops Prod.* **2018**, *117*, 224–234. [CrossRef]

60. Khaki, S.; Wang, L.; Archontoulis, S.V. A CNN-RNN Framework for Crop Yield Prediction. *Front. Plant Sci.* **2020**, *10*, 1750. [CrossRef] [PubMed]

61. Jiang, H.; Hu, H.; Zhong, R.; Xu, J.; Xu, J.; Huang, J.; Wang, S.; Ying, Y.; Lin, T. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: A case study of the US Corn Belt at the county level. *Glob. Chang. Biol.* **2020**, *26*, 1754–1766. [CrossRef] [PubMed]

62. Bornn, L.; Zidek, J.V. Efficient stabilization of crop yield prediction in the Canadian Prairies. *Agric. For. Meteorol.* **2012**, *152*, 223–232. [CrossRef]

63. Dai, X.; Huo, Z.; Wang, H. Simulation for response of crop yield to soil moisture and salinity with artificial neural network. *Field Crop. Res.* **2011**, *121*, 441–449. [CrossRef]

64. Farjam, A.; Omid, M.; Akram, A.; Fazel Niari, Z. A neural network based modeling and sensitivity analysis of energy inputs for predicting seed and grain corn yields. *J. Agric. Sci. Technol.* **2014**, *16*, 767–778.

65. Piekutowska, M.; Niedbała, G. Application of artificial neural networks for the prediction of quality characteristics of potato tubers—Innovator variety. *J. Res. Appl. Agric. Eng.* **2018**, *63*, 132–138.

66. Maya Gopal, P.S.; Bhargavi, R. A novel approach for efficient crop yield prediction. *Comput. Electron. Agric.* **2019**, *165*, 104968. [CrossRef]

67. Emamgholizadeh, S.; Parsaeian, M.; Baradaran, M. Seed yield prediction of sesame using artificial neural network. *Eur. J. Agron.* **2015**, *68*, 89–96. [CrossRef]

68. Torkashvand, A.M.; Ahmadi, A.; Nikravesh, N.L. Prediction of kiwifruit firmness using fruit mineral nutrient concentration by artificial neural network (ANN) and multiple linear regressions (MLR). *J. Integr. Agric.* **2017**, *16*, 1634–1644. [CrossRef]

69. Niedbała, G. Simple model based on artificial neural network for early prediction and simulation winter rapeseed yield. *J. Integr. Agric.* **2019**, *18*, 54–61. [CrossRef]

70. Ayoubi, S.; Sahrawat, K.L. Comparing multivariate regression and artificial neural network to predict barley production from soil characteristics in northern Iran. *Arch. Agron. Soil Sci.* **2011**, *57*, 549–565. [CrossRef]

71. 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.

72. Niedbała, G. Application of artificial neural networks for multi-criteria yield prediction ofwinter rapeseed. *Sustainability* **2019**, *11*, 533. [CrossRef]

73. Clevers, J.G.P.W.; Kooistra, L.; van den Brande, M.M.M. Using Sentinel-2 data for retrieving LAI and leaf and canopy chlorophyll content of a potato crop. *Remote Sens.* **2017**, *9*, 405. [CrossRef]

74. Cai, Y.; Guan, K.; Lobell, D.; Potgieter, A.B.; Wang, S.; Peng, J.; Xu, T.; Asseng, S.; Zhang, Y.; You, L.; et al. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* **2019**, *274*, 144–159. [CrossRef]

75. Piao, S.; Ciais, P.; Huang, Y.; Shen, Z.; Peng, S.; Li, J.; Zhou, L.; Liu, H.; Ma, Y.; Ding, Y.; et al. The impacts of climate change on water resources and agriculture in China. *Nature* **2010**, *467*, 43–51. [CrossRef] [PubMed]

76. Hara, P.; Szparaga, A. Ecological methods used to control fungi that cause diseases of the crop plant. *Rocz. Ochr. Sr.* **2018**, *20*, 1764–1775.

77. Lobell, D.B.; Field, C.B. Global scale climate–crop yield relationships and the impacts of recent warming. *Environ. Res. Lett.* **2007**, *2*, 014002. [CrossRef]

78. Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]

79. Jiang, Z.; Liu, C.; Ganapathysubramanian, B.; Hayes, D.J.; Sarkar, S. Predicting county-scale maize yields with publicly available data. *Sci. Rep.* **2020**, *10*, 14957. [CrossRef] [PubMed]

80. Tao, F.; Xiao, D.; Zhang, S.; Zhang, Z.; Rötter, R.P. Wheat yield benefited from increases in minimum temperature in the Huang-Huai-Hai Plain of China in the past three decades. *Agric. For. Meteorol.* **2017**, *239*, 1–14. [CrossRef]

81. Skrzyczyńska, J.; Gąsiorowska, B. *Uprawa roślin*; Uniwersytet Przyrodniczy we Wrocławiu: Wrocław, Poland; 2020; pp. 49–210.

82. Slafer, G.A.; Savin, R. Developmental Base Temperature in Different Phenological Phases of Wheat (Triticum aestivum). *J. Exp. Bot.* **1991**, *42*, 1077–1082. [CrossRef]

83. Tsimba, R.; Edmeades, G.O.; Millner, J.P.; Kemp, P.D. The effect of planting date on maize grain yields and yield components. *F. Crop. Res.* **2013**, *150*, 135–144. [CrossRef]

84. Rumpf, S.B.; Semenchuk, P.R.; Dullinger, S.; Cooper, E.J. Idiosyncratic Responses of High Arctic Plants to Changing Snow Regimes. *PLoS ONE* **2014**, *9*, e86281. [CrossRef]

85. Asseng, S.; Foster, I.; Turner, N.C. The impact of temperature variability on wheat yields. *Glob. Chang. Biol.* **2011**, *17*, 997–1012. [CrossRef]

86. Siebert, S.; Webber, H.; Rezaei, E.E. Weather impacts on crop yields—Searching for simple answers to a complex problem. *Environ. Res. Lett.* **2017**, *12*. [CrossRef]

87. Kern, A.; Barcza, Z.; Marjanović, H.; Árendás, T.; Fodor, N.; Bónis, P.; Bognár, P.; Lichtenberger, J. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agric. For. Meteorol.* **2018**, *260–261*, 300–320. [CrossRef]

88. Han, J.; Zhang, Z.; Cao, J.; Luo, Y.; Zhang, L.; Li, Z.; Zhang, J. Prediction of Winter Wheat Yield Based on Multi-Source Data and Machine Learning in China. *Remote Sens.* **2020**, *12*, 236. [CrossRef]

89. Osakabe, Y.; Osakabe, K.; Shinozaki, K.; Tran, L.-S.P. Response of plants to water stress. *Front. Plant Sci.* **2014**, *5*, 86. [CrossRef] [PubMed]

90. Semenov, M.A.; Shewry, P.R. Modelling predicts that heat stress, not drought, will increase vulnerability of wheat in Europe. *Sci. Rep.* **2011**, *1*, 66. [CrossRef] [PubMed]

91. Tesfamariam, E.H.; Annandale, J.G.; Steyn, J.M. Water Stress Effects on Winter Canola Growth and Yield. *Agron. J.* **2010**, *102*, 658–666. [CrossRef]

92. Tielbörger, K.; Kadmon, R. Temporal environmental variation tips the balance between facilitation and interference in desert plants. *Ecology* **2000**, *81*, 1544–1553. [CrossRef]

93. Levine, J.M.; McEachern, A.K.; Cowan, C. Rainfall effects on rare annual plants. *J. Ecol.* **2008**, *96*, 795–806. [CrossRef]

94. 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.

95. Jin, Y.; Chen, B.; Lampinen, B.D.; Brown, P.H. Advancing Agricultural Production With Machine Learning Analytics: Yield Determinants for California's Almond Orchards. *Front. Plant Sci.* **2020**, *11*, 290. [CrossRef] [PubMed]

96. Kaul, M.; Hill, R.L.; Walthall, C. Artificial neural networks for corn and soybean yield prediction. *Agric. Syst.* **2005**, *85*, 1–18. [CrossRef]

97. Żmudzka, E. The Climatic Background of Agricultural Production in Poland (1951–2000). *Misc. Geogr.* **2004**, *11*, 127–137. [CrossRef]

98. Morozova, S.V.; Polyanskaya, E.A.; Kononova, N.K.; Denisov, K.E.; Poletaev, I.S. The study of the dependence of spring crops yield on the abiotic environmental factors using nonlinear interpolation. In Proceedings of the IOP Conference Series: Earth and Environmental Science, Irkutsk, Russian, 23–27 September 2019; Volume 381.

99. Selyaninov, G.T. About agricultural climate assessment. *Work. Agric. Meteorol.* **1928**, *20*, 165–177.

100. Paltasingh, K.R.; Goyari, P.; Mishra, R.K. Measuring weather impact on crop yield using aridity index: Evidence from Odisha. *Agric. Econ. Res. Rev.* **2012**, *25*, 205–216.

101. Babushkina, E.A.; Belokopytova, L.V.; Zhirnova, D.F.; Shah, S.K.; Kostyakova, T.V. Climatically driven yield variability of major crops in Khakassia (South Siberia). *Int. J. Biometeorol.* **2018**, *62*, 939–948. [CrossRef] [PubMed]

102. Feng, X.; Vico, G.; Porporato, A. On the effects of seasonality on soil water balance and plant growth. *Water Resour. Res.* **2012**, *48*. [CrossRef]

103. Rodríguez-Iturbe, I.; Porporato, A. *Ecohydrology of Water-Controlled Ecosystems: Soil Moisture and Plant Dynamics*; Cambridge University Press: Cambridge, UK, 2007.

104. Bastiaanssen, W.G.M.; Menenti, M.; Feddes, R.A.; Holtslag, A.A.M. A remote sensing surface energy balance algorithm for land (SEBAL). 1. Formulation. *J. Hydrol.* **1998**, *212–213*, 198–212. [CrossRef]

105. Allen, R.G.; Tasumi, M.; Trezza, R. Satellite-Based Energy Balance for Mapping Evapotranspiration with Internalized Calibration (METRIC)—Model. *J. Irrig. Drain. Eng.* **2007**, *133*, 380–394. [CrossRef]

106. Karimi, P.; Bastiaanssen, W.G.M.; Molden, D. Water Accounting Plus (WA+)—a water accounting procedure for complex river basins based on satellite measurements. *Hydrol. Earth Syst. Sci.* **2013**, *17*, 2459–2472. [CrossRef]

107. Mhawej, M.; Caiserman, A.; Nasrallah, A.; Dawi, A.; Bachour, R.; Faour, G. Automated evapotranspiration retrieval model with missing soil-related datasets: The proposal of SEBALI. *Agric. Water Manag.* **2020**, *229*, 105938. [CrossRef]

108. Asgarzadeh, H.; Mosaddeghi, M.R.; Mahboubi, A.A.; Nosrati, A.; Dexter, A.R. Soil water availability for plants as quantified by conventional available water, least limiting water range and integral water capacity. *Plant Soil* **2010**, *335*, 229–244. [CrossRef]

109. Kitchen, N.R.; Drummond, S.T.; Lund, E.D.; Sudduth, K.A.; Buchleiter, G.W. Soil Electrical Conductivity and Topography Related to Yield for Three Contrasting Soil–Crop Systems. *Agron. J.* **2003**, *95*, 483–495. [CrossRef]

110. Manrique, L.A.; Jones, C.A.; Dyke, P.T. Predicting Cation-Exchange Capacity from Soil Physical and Chemical Properties. *Soil Sci. Soc. Am. J.* **1991**, *55*, 787–794. [CrossRef]

111. Saikh, H.; Varadachari, C.; Ghosh, K. Effects of deforestation and cultivation on soil CEC and contents of exchangeable bases: A case study in Simlipal National Park, India. *Plant Soil* **1998**, *204*, 175–181. [CrossRef]

112. Pernes-Debuyser, A.; Tessier, D. Soil physical properties affected by long-term fertilization. *Eur. J. Soil Sci.* **2004**, *55*, 505–512. [CrossRef]

113. Crane-Droesch, A. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.* **2018**, *13*, 114003. [CrossRef]

114. Adisa, O.M.; Botai, C.M.; Botai, J.O.; Hassen, A.; Darkey, D.; Tesfamariam, E.; Adisa, A.F.; Adeola, A.M.; Ncongwane, K.P. Analysis of agro-climatic parameters and their influence on maize production in South Africa. *Theor. Appl. Climatol.* **2018**, *134*, 991–1004. [CrossRef]

115. Amaratunga, V.; Wickramasinghe, L.; Perera, A.; Jayasinghe, J.; Rathnayake, U. Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data. *Math. Probl. Eng.* **2020**, *2020*, 1–11. [CrossRef]

116. Guo, Y.; Xiang, H.; Li, Z.; Ma, F.; Du, C. Prediction of Rice Yield in East China Based on Climate and Agronomic Traits Data Using Artificial Neural Networks and Partial Least Squares Regression. *Agronomy* **2021**, *11*, 282. [CrossRef]

117. Zhang, L.; Zhang, J.; Kyei-boahen, S.; Zhang, M. Simulation and Prediction of Soybean Growth and Development under Field Conditions. *Am. J.Agric. Environ. Sci* **2010**, *7*, 374–385.

118. Niedbała, G.; Piekutowska, M.; Weres, J.; Korzeniewicz, R.; Witaszek, K.; Adamski, M.; Pilarski, K.; Czechowska-Kosacka, A.; Krysztofiak-Kaniewska, A. Application of artificial neural networks for yield modeling of winter rapeseed based on combined quantitative and qualitative data. *Agronomy* **2019**, *9*, 781. [CrossRef]

119. Fageria, N.K.; Filho, M.P.B.; Moreira, A.; Guimarães, C.M. Foliar Fertilization of Crop Plants. *J. Plant Nutr.* **2009**, *32*, 1044–1064. [CrossRef]

120. Haytova, D. A Review of Foliar Fertilization of Some Vegetables Crops. *Annu. Rev. Res. Biol.* **2013**, *3*, 455–465.

121. Dordas, C. Role of nutrients in controlling plant diseases in sustainable agriculture. A review. *Agron. Sustain. Dev.* **2008**, *28*, 33–46. [CrossRef]

122. Hirel, B.; Tétu, T.; Lea, P.J.; Dubois, F. Improving Nitrogen Use Efficiency in Crops for Sustainable Agriculture. *Sustainability* **2011**, *3*, 1452–1485. [CrossRef]

123. Kotsiantis, S.B. *Supervised Machine Learning: A Review of Classification Techniques*; IOS Press: Amsterdam, The Netherlands, 2007; pp. 3–24.

124. Gibert, K.; Sànchez-Marrè, M.; Izquierdo, J. A survey on pre-processing techniques: Relevant issues in the context of environmental data mining. *AI Commun.* **2016**, *29*, 627–663. [CrossRef]

125. Lillesand, T.M.; Kiefer, R.W. *Remote Sensing and Image Interpretation*; John Wiley&Sons: New York, NY, USA, 1994.

126. Basso, B.; Cammarano, D.; Carfagna, E. Review of Crop Yield Forecasting Methods and Early Warning Systems. *First Meet. Sci. Advis. Comm. Glob. Strateg. Improv. Agric. Rural Stat.* **2013**, *41*, 1–56.

127. Yu, L.; Liang, L.; Wang, J.; Zhao, Y.; Cheng, Q.; Hu, L.; Liu, S.; Yu, L.; Wang, X.; Zhu, P.; et al. Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *Int. J. Remote Sens.* **2014**, *35*, 4573–4588. [CrossRef]

128. Yu, L.; Xu, Y.; Xue, Y.; Li, X.; Cheng, Y.; Liu, X.; Porwal, A.; Holden, E.-J.; Yang, J.; Gong, P. Monitoring surface mining belts using multiple remote sensing datasets: A global perspective. *Ore Geol. Rev.* **2018**, *101*, 675–687. [CrossRef]

129. Jin, Y. Monitoring forage production in rangeland using remote sensing observations. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 3555–3556.

130. Chen, B.; Huang, B.; Xu, B. Multi-source remotely sensed data fusion for improving land cover classification. *ISPRS J. Photogramm. Remote Sens.* **2017**, *124*, 27–39. [CrossRef]

131. Chen, B.; Jin, Y.; Brown, P. An enhanced bloom index for quantifying floral phenology using multi-scale remote sensing observations. *ISPRS J. Photogramm. Remote Sens.* **2019**, *156*, 108–120. [CrossRef]

132. Gómez, S.; Sanz, C. Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data. *Remote Sens.* **2019**, *11*, 1745. [CrossRef]

133. Chen, P.-Y.; Fedosejevs, G.; Tiscareño-LóPez, M.; Arnold, J.G. Assessment of MODIS-EVI, MODIS-NDVI and VEGETATION-NDVI Composite Data Using Agricultural Measurements: An Example at Corn Fields in Western Mexico. *Environ. Monit. Assess.* **2006**, *119*, 69–82. [CrossRef]

134. Bala, S.K.; Islam, A.S. Correlation between potato yield and MODIS-derived vegetation indices. *Int. J. Remote Sens.* **2009**, *30*, 2491–2507. [CrossRef]

135. Miao, G.; Guan, K.; Yang, X.; Bernacchi, C.J.; Berry, J.A.; DeLucia, E.H.; Wu, J.; Moore, C.E.; Meacham, K.; Cai, Y.; et al. Sun-Induced Chlorophyll Fluorescence, Photosynthesis, and Light Use Efficiency of a Soybean Field from Seasonally Continuous Measurements. *J. Geophys. Res. Biogeosciences* **2018**, *123*, 610–623. [CrossRef]

136. Sruthi, S.; Aslam, M.A.M. Agricultural Drought Analysis Using the NDVI and Land Surface Temperature Data; a Case Study of Raichur District. *Aquat. Procedia* **2015**, *4*, 1258–1264. [CrossRef]

137. Galvão, L.S.; dos Santos, J.R.; Roberts, D.A.; Breunig, F.M.; Toomey, M.; de Moura, Y.M. On intra-annual EVI variability in the dry season of tropical forest: A case study with MODIS and hyperspectral data. *Remote Sens. Environ.* **2011**, *115*, 2350–2359. [CrossRef]

138. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E..; Gao, X.; Ferreira, L. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [CrossRef]

139. Fitzgerald, G.; Rodriguez, D.; O'Leary, G. Measuring and predicting canopy nitrogen nutrition in wheat using a spectral index—The canopy chlorophyll content index (CCCI). *Field Crop. Res.* **2010**, *116*, 318–324. [CrossRef]
140. Uno, Y.; Prasher, S.O.; Lacroix, R.; Goel, P.K.; Karimi, Y.; Viau, A.; Patel, R.M. Artificial neural networks to predict corn yield from Compact Airborne Spectrographic Imager data. *Comput. Electron. Agric.* **2005**, *47*, 149–161. [CrossRef]
141. Panda, S.S.; Ames, D.P.; Panigrahi, S. Application of Vegetation Indices for Agricultural Crop Yield Prediction Using Neural Network Techniques. *Remote Sens.* **2010**, *2*, 673–696. [CrossRef]
142. Fernandes, J.L.; Ebecken, N.F.F.; Esquerdo, J.C.D.M. Sugarcane yield prediction in Brazil using NDVI time series and neural networks ensemble. *Int. J. Remote Sens.* **2017**, *38*, 4631–4644. [CrossRef]
143. Chen, P.; Jing, Q. A comparison of two adaptive multivariate analysis methods (PLSR and ANN) for winter wheat yield forecasting using Landsat-8 OLI images. *Adv. Sp. Res.* **2017**, *59*, 987–995. [CrossRef]
144. Aghighi, H.; Azadbakht, M.; Ashourloo, D.; Shahrabi, H.S.; Radiom, S. Machine Learning Regression Techniques for the Silage Maize Yield Prediction Using Time-Series Images of Landsat 8 OLI. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 4563–4577. [CrossRef]
145. Kanning, M.; Kühling, I.; Trautz, D.; Jarmer, T. High-Resolution UAV-Based Hyperspectral Imagery for LAI and Chlorophyll Estimations from Wheat for Yield Prediction. *Remote Sens.* **2018**, *10*, 2000. [CrossRef]
146. Rahman, M.; Robson, A.; Bristow, M. Exploring the Potential of High Resolution WorldView-3 Imagery for Estimating Yield of Mango. *Remote Sens.* **2018**, *10*, 1866. [CrossRef]
147. Serele, C.Z.; Gwyn, Q.H.J.; Boisvert, J.B.; Pattey, E.; McLaughlin, N.; Daoust, G. Corn yield prediction with artificial neural network trained using airborne remote sensing and topographic data. In Proceedings of the IGARSS 2000. IEEE 2000 International Geoscience and Remote Sensing Symposium, Taking the Pulse of the Planet: The Role of Remote Sensing in Managing the Environment, Piscataway, NJ, USA, 24–28 July 2000; Volume 1, pp. 384–386.
148. Feng, L.; Zhang, Z.; Ma, Y.; Du, Q.; Williams, P.; Drewry, J.; Luck, B. Alfalfa Yield Prediction Using UAV-Based Hyperspectral Imagery and Ensemble Learning. *Remote Sens.* **2020**, *12*, 2028. [CrossRef]
149. Al-Gaadi, K.A.; Hassaballa, A.A.; Tola, E.; Kayad, A.G.; Madugundu, R.; Alblewi, B.; Assiri, F. Prediction of Potato Crop Yield Using Precision Agriculture Techniques. *PLoS ONE* **2016**, *11*, e0162219. [CrossRef]
150. Evrendilek, F.; Gulbeyaz, O. Deriving Vegetation Dynamics of Natural Terrestrial Ecosystems from MODIS NDVI/EVI Data over Turkey. *Sensors* **2008**, *8*, 5270–5302. [CrossRef]
151. Kozłowski, R.J.; Kozłowski, J.; Przybył, K.; Niedbała, G.; Mueller, W.; Okoń, P.; Wojcieszak, D.; Koszela, K.; Kujawa, S. Image analysis techniques in the study of slug behaviour. *SPIE* **2016**, *10033*, 100332I. [CrossRef]
152. Gahegan, M.; Ehlers, M. A framework for the modelling of uncertainty between remote sensing and geographic information systems. *ISPRS J. Photogramm. Remote Sens.* **2000**, *55*, 176–188. [CrossRef]
153. Rocchini, D.; Foody, G.M.; Nagendra, H.; Ricotta, C.; Anand, M.; He, K.S.; Amici, V.; Kleinschmit, B.; Förster, M.; Schmidtlein, S.; et al. Uncertainty in ecosystem mapping by remote sensing. *Comput. Geosci.* **2013**, *50*, 128–135. [CrossRef]
154. Rawte, V.; Anuradha, G. Fraud detection in health insurance using data mining techniques. In Proceedings of the 2015 International Conference on Communication, Information & Computing Technology (ICCICT), Mumbai, India, 15–17 January 2015; pp. 1–5.
155. Matsushita, B.; Yang, W.; Chen, J.; Onda, Y.; Qiu, G. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. *Sensors* **2007**, *7*, 2636–2651. [CrossRef] [PubMed]
156. Son, N.T.; Chen, C.F.; Chen, C.R.; Minh, V.Q.; Trung, N.H. A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. *Agric. For. Meteorol.* **2014**, *197*, 52–64. [CrossRef]
157. Bolton, D.K.; Friedl, M.A. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* **2013**, *173*, 74–84. [CrossRef]
158. HOLBEN, B.; JUSTICE, C. An examination of spectral band ratioing to reduce the topographic effect on remotely sensed data. *Int. J. Remote Sens.* **1981**, *2*, 115–133. [CrossRef]
159. Wang, Q.; Adiku, S.; Tenhunen, J.; Granier, A. On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sens. Environ.* **2005**, *94*, 244–255. [CrossRef]
160. Zhang, X.; Zhang, Q. Monitoring interannual variation in global crop yield using long-term AVHRR and MODIS observations. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 191–205. [CrossRef]
161. Benecki, P.; Kawulok, M.; Kostrzewa, D.; Skonieczny, L. Evaluating super-resolution reconstruction of satellite images. *Acta Astronaut.* **2018**, *153*, 15–25. [CrossRef]
162. Goward, S.N.; Markham, B.; Dye, D.G.; Dulaney, W.; Yang, J. Normalized difference vegetation index measurements from the advanced very high resolution radiometer. *Remote Sens. Environ.* **1991**, *35*, 257–277. [CrossRef]
163. NAGARAJA RAO, C.R.; CHEN, J. Post-launch calibration of the visible and near-infrared channels of the Advanced Very High Resolution Radiometer on the NOAA-14 spacecraft. *Int. J. Remote Sens.* **1996**, *17*, 2743–2747. [CrossRef]
164. Trishchenko, A.P. Effects of spectral response function on surface reflectance and NDVI measured with moderate resolution satellite sensors: Extension to AVHRR NOAA-17, 18 and METOP-A. *Remote Sens. Environ.* **2009**, *113*, 335–341. [CrossRef]
165. Albarakat, R.; Lakshmi, V. Comparison of Normalized Difference Vegetation Index Derived from Landsat, MODIS, and AVHRR for the Mesopotamian Marshes Between 2002 and 2018. *Remote Sens.* **2019**, *11*, 1245. [CrossRef]

166. Gitelson, A.A.; Kaufman, Y.J. MODIS NDVI Optimization To Fit the AVHRR Data Series—Spectral Considerations. *Remote Sens. Environ.* **1998**, *66*, 343–350. [CrossRef]

167. Li, A.; Liang, S.; Wang, A.; Qin, J. Estimating Crop Yield from Multi-temporal Satellite Data Using Multivariate Regression and Neural Network Techniques. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 1149–1157. [CrossRef]

168. Mkhabela, M.S.; Bullock, P.; Raj, S.; Wang, S.; Yang, Y. Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. For. Meteorol.* **2011**, *151*, 385–393. [CrossRef]

169. Chen, B.; Li, J.; Jin, Y. Deep Learning for Feature-Level Data Fusion: Higher Resolution Reconstruction of Historical Landsat Archive. *Remote Sens.* **2021**, *13*, 167. [CrossRef]

170. 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]

171. Zhang, L.; Traore, S.; Ge, J.; Li, Y.; Wang, S.; Zhu, G.; Cui, Y.; Fipps, G. Using boosted tree regression and artificial neural networks to forecast upland rice yield under climate change in Sahel. *Comput. Electron. Agric.* **2019**, *166*, 105031. [CrossRef]

172. Khaki, S.; Wang, L. Crop Yield Prediction Using Deep Neural Networks. *Front. Plant Sci.* **2019**, *10*, 621. [CrossRef]

173. Zhang, M.; Zhang, Y.; Vo, D.T. Gated neural networks for targeted sentiment analysis. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 3087–3093.

174. Kuwata, K.; Shibasaki, R. Estimating corn yield in the united states with modis evi and machine learning methods. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *III–8*, 131–136. [CrossRef]

175. Magney, T.S.; Frankenberg, C.; Fisher, J.B.; Sun, Y.; North, G.B.; Davis, T.S.; Kornfeld, A.; Siebke, K. Connecting active to passive fluorescence with photosynthesis: A method for evaluating remote sensing measurements of Chl fluorescence. *New Phytol.* **2017**, *215*, 1594–1608. [CrossRef]