

## Article

# Data Mining and Machine Learning Algorithms for Optimizing Maize Yield Forecasting in Central Europe

Endre Harsányi <sup>1,2</sup>, Bashar Bashir <sup>3</sup> , Sana Arshad <sup>4</sup> , Akasairi Ocwa <sup>1,5</sup> , Attila Vad <sup>2</sup>, Abdullah Alsalman <sup>3</sup>, István Bácskai <sup>2</sup>, Tamás Rátonyi <sup>1</sup>, Omar Hijazi <sup>6</sup>, Adrienn Széles <sup>1</sup>  and Safwan Mohammed <sup>1,2,\*</sup> 

<sup>1</sup> Institute of Land Use, Technical and Precision Technology, Faculty of Agricultural and Food Sciences and Environmental Management, University of Debrecen, 4032 Debrecen, Hungary

<sup>2</sup> Institutes for Agricultural Research and Educational Farm, University of Debrecen, Böszörményi 138, 4032 Debrecen, Hungary

<sup>3</sup> Department of Civil Engineering, College of Engineering, King Saud University, P.O. Box 800, Riyadh 11421, Saudi Arabia

<sup>4</sup> Department of Geography, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan

<sup>5</sup> Department of Agriculture Production, Faculty of Agriculture, Kyambogo University, Kyambogo, Kampala P.O. Box 1, Uganda

<sup>6</sup> Chair of Wood Science, Technical University of Munich, 85354 Freising, Germany

\* Correspondence: safwan@agr.unideb.hu

**Abstract:** Artificial intelligence, specifically machine learning (ML), serves as a valuable tool for decision support in crop management under ongoing climate change. However, ML implementation to predict maize yield is still limited in Central Europe, especially in Hungary. In this context, we assessed the performance of four ML algorithms (Bagging (BG), Decision Tree (DT), Random Forest (RF) and Artificial Neural Network-Multi Layer Perceptron (ANN-MLP)) in predicting maize yield based on four different input scenarios. The collected data included both agricultural data (production (PROD) (ton) and maize cropped area (AREA) (ha)) and climate data (annual mean temperature °C (Tmean), precipitation (PRCP) (mm), rainy days (RD), frosty days (FD) and hot days (HD)). This research adopted four scenarios, as follows: SC1: AREA+ PROD+ Tmean+ PRCP+ RD+ FD+ HD; SC2: AREA+ PROD; SC3: Tmean+ PRCP+ RD+ FD+ HD; and SC4: AREA+ PROD+ Tmean+ PRCP. In the training stage, ANN-MLP-SC1 and ANN-MLP-SC4 outperformed other ML algorithms; the correlation coefficient (r) was 0.99 for both, while the root mean squared errors (RMSEs) were 107.9 (ANN-MLP-SC1) and 110.7 (ANN-MLP-SC4). In the testing phase, the ANN-MLP-SC4 had the highest r value (0.96), followed by ANN-MLP-SC1 (0.94) and RF-SC2 (0.94). The 10-fold cross validation also revealed that the ANN-MLP-SC4 and ANN-MLP-SC1 have the highest performance. We further evaluated the performance of the ANN-MLP-SC4 in predicting maize yield on a regional scale (Budapest). The ANN-MLP-SC4 succeeded in reaching a high-performance standard (r = 0.98, relative absolute error = 21.87%, root relative squared error = 20.4399% and RMSE = 423.23). This research promotes the use of ANN as an efficient tool for predicting maize yield, which could be highly beneficial for planners and decision makers in developing sustainable plans for crop management.

**Keywords:** maize yield; climate; multilayer perceptron; random forest; optimum model



**Citation:** Harsányi, E.; Bashir, B.; Arshad, S.; Ocwa, A.; Vad, A.; Alsalman, A.; Bácskai, I.; Rátonyi, T.; Hijazi, O.; Széles, A.; et al. Data Mining and Machine Learning Algorithms for Optimizing Maize Yield Forecasting in Central Europe. *Agronomy* **2023**, *13*, 1297. <https://doi.org/10.3390/agronomy13051297>

Academic Editor: Xiuliang Jin, Hao Yang, Zhenhai Li, Changping Huang and Dameng Yin

Received: 14 March 2023

Revised: 22 April 2023

Accepted: 30 April 2023

Published: 4 May 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Cereal crops, mainly wheat, maize and rice, contribute to sustainable livelihoods and food security [1]. In fact, staple cereal crops are projected to play an indispensable role for food security to ensure sufficient and affordable protein and calorie intakes in diets [2]. Maize is the second most widely grown crop next to wheat and amounts to 197 million hectares of global land area with substantial sharing by Sub-Saharan Africa, Latin America and Asia [3]. Unlike other cereal crops, maize is a versatile crop feeding both humans and livestock and it is also utilized in the bio-energy sector of industrial

processing [4]. The global annual maize production was recently (2016 to 2018) estimated at 1127 million tons [5]. The global maize yield is increasing annually by 1.6%, but it is negatively impacted by climate change and is insufficient to meet the demands of a growing population [4,6]. According to Okolie, et al. [7], the nexus between climate change and crop yields is crop and region-specific and contingent on variances in soil, management, time and length of crop exposure to numerous climate variables. It is reported that by 2050, maize production is expected to decline by 10% and 9–19% in sub-Saharan Africa and South Asia, respectively [8]. Furthermore, it is expected that global climatic change will reduce maize yield by 60%, wheat by 20%, rice by 35% and sorghum by 50% [7].

Europe contributes around 35% of global wheat production [9] and 11% of the world's maize [3]. Several biotic (diseases and insects) and abiotic (droughts and heatwaves) stresses have led to a shift in maize mega-environments and the introduction of new breeding varieties [10]. Climatic extremes, such as high temperatures, increase the respiration rate and reduce the grain filling in cereal crops, causing a rapid decline in yield [11], depending on the intensity and frequency of the changes in the meteorological variables [11–13]. Due to the uncertainty of potential losses in crop yields caused by climate change, it is essential to forecast future changes and develop strategies to prevent such losses [14]. Pant, et al. [15] highlighted that predicting crop yield is a critical challenge in crop production because it relies on various weather factors, including rainfall, humidity and temperature, which are subject to fluctuations. To address this, data mining and the application of artificial intelligence is widely adopted.

Data mining accumulates undetected evidence from vast databases and assists in the detailed analysis of prospective patterns, facilitating informed decisions. Artificial intelligence, specifically ML, serves as a valuable tool for decision support in crop management. By analyzing vast amounts of data on crop growth, weather patterns, soil conditions and other relevant factors, “ML” algorithms can identify patterns and make predictions that can inform crop management decisions [16]. Machine learning has higher flexibility and provides sound, faster predictions compared to simulation crop modelling [17]. Several machine and deep learning algorithms such as Random Forest (RF), Support Vector Machines (SVM) and Neural Networks (NN), are used for crop yield predictions in varying environmental conditions [18–21]. Support Vector Machines have the ability to make robust predictions if data sets have several attributes, but their functionality may have hardware restrictions. On the other hand, random forest algorithms have high classifier and superior predictions compared to multiple linear regression models [22]. According to [23], simple linear and generalized linear (GLM) models have restrictive assumptions with the normal distribution trained on the value of predictors with a constant variance. Generally, decision tree ensemble learning methods—the Random Forest Regressor and Gradient Boosting Regressor—have higher accuracy [15,24]. In China, ref. [25] revealed that the highest estimation accuracy was in the Random Forest Regression. In India, the decision tree produced the highest (96%) accuracy of crop yield prediction [15]. The prediction effectiveness of machine learning algorithms is assessed based on the models' abilities to reduce the bias, the variance or both; hence, informing the basis for recommending the use of a particular algorithm. For instance, if prediction with a low error is the aim, then weighted ensemble models are selected; if detecting the correct forecast direction is required, then a stacked LASSO regression is chosen [18,26]. An overview of ML algorithms used in crop production in previous studies are presented in Table 1.

The non-unidirectional performance of different algorithms in Table 1 above raises this question: which algorithm will provide the precise prediction of maize yields in Central Europe? Providing answers to this question requires testing the performance of different models. In this study, we used the following climate elements to predict the yield of maize: the mean temperature, precipitation, number of frosty days, number of hot days and number of rainy days were incorporated into the model hyper parameterization. The specific objectives of this study were to: (I) analyze the changes in chosen climate parameters along with maize production elements across Hungary between 1921 to 2018;

(II) to assess the performance of four ML algorithms (Bagging (BG), Decision Table (DT), Random Forest (RF) and Artificial Neural Network-Multi Layer Perceptron (ANN-MLP)) in predicting maize yield depending on four scenarios; and (III) to test the performance of the best combination of ML algorithms and scenarios for predicting maize yield in a regional scale across Hungary.

**Table 1.** The application of machine learning in predicting the growth and yield of maize and other crops.

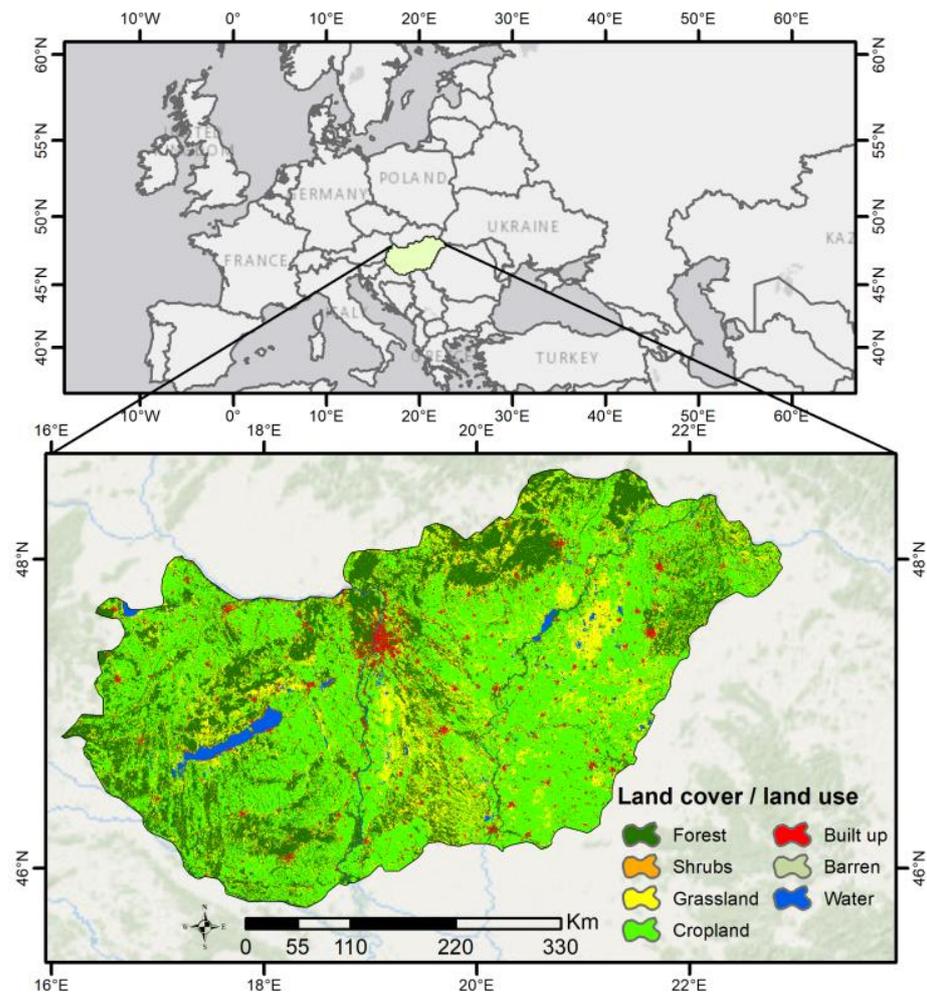
Country	Reference	Crop	Method	Comment
Spain	[21]	Olive crop	GLM algorithm, Support Vector Machines (SVM), Gaussian and Linear Kernel	Yield predicted with low error. Parameters that introduce noise in the model must be discarded.
India	[27]		Random Forest Algorithm	Superior prediction compared to Decision Tree.
USA	[17]	Corn	Random forest, XGBoost and LightGBM, Stacked Generalization	Models could not predict better than base learners due to blocked sequential procedure.
India	[15]	Maize, wheat, rice, potatoes	Decision Tree Regressor, Gradient Boosting Regressor, Random Forest Regressor, SVM	Decision Tree Regressor had the highest (96%) prediction accuracy.
Ethiopia, Kenya, Tanzania, Malawi and Mozambique.	[28]	Maize	Linear Algorithms (Logistic Regression (LR)) and Linear Discriminant Analysis (LDA). Nonlinear algorithms [K-Nearest Neighbor (KNN), Classification and Regression Trees (CART), Gaussian Naive Bayes (NB) and Support Vector Machine (SVM)].	Support Vector Machine was the worst algorithm.
China	[25]	Maize	Random Forest Regression (RFR), Gradient Boosting Decision Tree (GBDT)	RFR had high yield estimation accuracy.
USA	[18]	Maize	Random Forest, XGBoost, Optimal Ensemble, Benchmark Ensemble, Linear and Ridge Regression	Had high accuracy.
Bulgaria, Germany, Spain, France, Hungary Italy, the Netherlands, Poland, Romania	[29]	Wheat, barley, sunflower, grain, maize, sugar beets, potato	Ridge Regression, K-Nearest Neighbors (KNN) Regression, Support Vector Machines Regression (SVR), Gradient Boosted Decision Trees (GBDT)	Machine learning forecasts had lower uncertainty than a trend model.

## 2. Materials and Methods

### 2.1. Study Area

The study was conducted in a significant agricultural country of Central Europe lying between 45°55' N–48°60' N and 16°10' E–22°50' E and occupying a total area of 93,000 km<sup>2</sup>. The country's landscape is dominated by the Carpathian Mountains in the north and the Great Hungarian Plain in the east. The Danube River flows through the country, dividing it into two main regions: Transdanubia to the west and the Great Hungarian Plain to the east. The country has a mild continental climate with warm summers and cold winters [30], and its primary land use consists of forests, grasslands, cropland, built-up areas and water bodies (Figure 1). The land-use structure of the country follows a specific pattern imposed

by the natural landscape, e.g., the forest area is predominantly occupied in mountainous and hilly terrains, while great plains of the country are mostly cultivated with winter wheat, maize, sunflower and rapeseed, which approximates 67% of the arable land [31]. Among the major crops, the area of maize ranks second after wheat, with a significant rise in yield in recent decades. The phenological stages of maize include sowing–canopy expansion (April–June), flowering–grain filling (July–August) and ripening–harvesting (September–October) [32].



**Figure 1.** Location and land cover or land-use distribution in the study area (Hungary).

## 2.2. The Data Set

A historical time-series of 98 years (1921–2018) was used for maize yield prediction in Hungary using multisource data. Our study utilized two types of datasets as explanatory variables for maize yield prediction (response variable). First, the agriculture data: this included maize production (PROD) (ton) and the maize cropped area (AREA) (ha). Second, climate data: this was employed for yield predictions and comprised the annual mean temperature °C (Tmean), precipitation (PRCP), rainy days (RD), frosty days (FD) and hot days (HD). Due to the limited availability of all climatic variables on a seasonal scale, data was collected for Hungary as a sample station on a mean annual basis from the website, Meteorological data of Hungary ([ksh.hu](http://ksh.hu)), (accessed on 1 April 2021) published and updated by the Hungarian central statistical office. The agricultural input variables were also obtained from the Hungarian central statistical office (Table 2).

**Table 2.** Data sets used in the study (accessed on 1 April 2021).

Input Data	Variables Selected	Time Period	Source
Agriculture	Cropped area (hectares), Crop production (tons), Crop yield (tons/ha)	1921–2018	<a href="https://www.ksh.hu/agricultural_census_long_time_series">https://www.ksh.hu/agricultural_census_long_time_series</a>
Climate	Mean temperature °C (Tmean), precipitation mm (PRCP), Rainy days (RD), frosty days (FD), heat days (HD)	1921–2018	<a href="https://www.ksh.hu/stadat_files/kor/en/kor0037.html">https://www.ksh.hu/stadat_files/kor/en/kor0037.html</a>

The climatic variables selected for the yield forecasting have a strong impact on crop yields causing yield gains or losses associated with precipitation, high mean temperatures and drought conditions [30,31,33]. Furthermore, technological improvements have facilitated an increase in the cropped area and in production. Hence, the selected parameters in our study are significant for historical maize yield forecasting, which employs ML methods. All input variables for yield forecasting were collected on the same temporal scale with no missing data, which provided a homogeneity with limited uncertainties in prediction modelling.

### 2.3. Trend (Mann–Kendall, Sen’s Slope and Sequential Mann–Kendall) and Correlation Analysis

The temporal and linear relationships between predictors (climatic and agricultural input variables) of the maize crop yield are examined using a trend analysis and Pearson correlation. The widely used non-parametric Mann–Kendall trend test [34,35], along with Sen’s slope (assuming no autocorrelation in time-series), examines the long-term monotonic trends in the explanatory and response variables. The trend is analyzed based on the null hypothesis  $H_0$  (assuming no significant trend) and alternate hypothesis  $H_1$  (assuming a significant trend) at a significance level of  $\alpha = 0.05$ . Moreover, to investigate the trend changing points (in current years), also called the intersection of breaking points, the Sequential Mann–Kendall (SQMK) is also applied [36]. To support the findings of the trend analysis, a polynomial fit is developed on a historical time-series of all input variables in Origin Pro 2022.

Moreover, before predicting the maize yield employing non-linear ML algorithms, it is important to examine the linear relationship between explanatory and response variables. The Pearson correlation is also employed in various ML studies as an exploratory data analysis of predictors and response variables [37,38]. Hence, we applied the Pearson correlation to examine the significance ( $p < 0.001$ ) of the linear relationships between all climatic predictors (mean temperature (Tmean), precipitation (PRCP), hot days (HD), frosty days (FD) and rainy days (RD)), crop-related predictors (maize cropped area (AREA) and maize production (PROD)) and the response variable, i.e., yield. Overall, the trend and correlation analyses also supported a more accurate interpretation of the findings of the ML methods.

### 2.4. The Methodology and Machine Learning Models Used for Maize Yield Prediction

We predict maize yield in Hungary from input data based on four scenarios (Table 3) by employing four significant machine learning approaches including Bagging (BG), Decision Table (DT), Random Forest (RF) and Artificial Neural Network-Multi Layer Perceptron (ANN-MLP). Algorithms are employed in an open-source data mining GUI, i.e., WEKA, which provides various data mining and machine learning opportunities to solve classification and regression problems [39]. For better optimization of yield forecasting, four scenarios (SC1–SC4) are developed with different combinations of explanatory variables. SC1 utilizes all climatic (Tmean+ PRCP+ RD+ FD+ HD) and agricultural (AREA+ PROD) variables, while SC2 utilizes only the agricultural (AREA+ PROD) input variables to predict maize yield. SC3 utilizes only climatic (Tmean+ PRCP+ RD+ FD+ HD) variables and SC4 utilizes agricultural (AREA+ PROD) and two climatic (Tmean+ PRCP) input variables

for predicting maize yield (Table 3). The main reason for developing four scenarios to forecast maize yield was to explore and identify the most significant climatic predictors in combination with the agricultural inputs. All the ML algorithms are applied in every scenario to find the optimum model and scenario combinations. To perform ML, the dataset ( $n = 98$ ) is randomly split into 80% ( $n = 78$ ) for training and 20% ( $n = 20$ ) for testing. Furthermore, a 10-fold cross validation was used on the entire dataset with the batch size and iteration equaling 100. This provides a robust estimate of the model's performance, with accurate estimation and generalization, and helped to reduce the impact of random sampling variation, therefore ensuring the better performance of the ML model. The details of each ML algorithm with the specified hyperparameters are presented in Table 4.

**Table 3.** Different scenarios based on different input variables for predicting maize yield.

Scenario	Input *	Output	Photoprint of the Scenario
SC1	AREA+ PROD+ Tmean+ PRCP+ RD+ FD+ HD	Maize yield	Agricultural data+ climate data
SC2	AREA+ PROD	Maize yield	Agricultural data
SC3	Tmean+ PRCP+ RD+ FD+ HD	Maize yield	Climate data
SC4	AREA+ PROD+ Tmean+ PRCP	Maize yield	Agricultural data+ climate data

\* Tmean: mean temperature, PRCP: precipitation, HD: hot days, FD: frosty days, RD: rainy days, AREA: maize cropped area, PROD: maize production.

**Table 4.** Parameter selection of all machine learning algorithms used in the present study.

Algorithm	Parameters
Bagging	Algorithm = 'weka.classifier.meta.bagging', pool size P = 100, seed S = 1, num-slot = 1, Base learner = weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0, iteration = 10
Decision Table	Algorithm = weka.classifiers.rules.DecisionTable, number of rules R = 10, CV = leave one out, Search method S = Best First, search direction D = forward, start set = no attributes, subsets evaluated = 31
Random Forest	weka.classifiers.trees.RandomForest, ntrees I = 100, num-slot = 1, max tree depth = 0, variance V = 0.001, seed S = 1, batch-size = 100, classifier capabilities = -do-not-check-capabilities
ANN-MLP	Learning rate L = 0.3, momentum M = 0.2, Activation function = sigmoid, Num of epochs to train = 500, E = threshold for consecutive errors = 20, Regularization = weight decay

#### 2.4.1. The Bagging Algorithm

The Bagging (BG) algorithm, also called Bootstrap aggregation, is a non-parametric and powerful machine learning technique that generates multiple ( $B$ ) separate training datasets using the bootstrap sampling method and trains a separate regression model on each subset [40]. Each subset is created by randomly selecting data instances with replacements. The model is trained on the  $b$ th bootstrapped training set to obtain  $\hat{f}_b(x)$ , and finally average all the predictions. It is presented by:

$$\hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x) \quad (1)$$

where  $\hat{f}_{bag}(x)$  is the final prediction obtained by averaging the predictions of  $B$  models trained on the different bootstrap samples of the data.

#### 2.4.2. The Decision Table Algorithm

The Decision Table (DT) algorithm is a precise method for predicting numerical outcomes using a sequence of "If-Then" rules. To create a decision table, all possible combinations of input variables are mapped to find the best set of input attributes that provide the most accurate prediction output [41]. When a new data item is presented, the algorithm assigns it to a category by matching its non-class values with the corresponding line in the decision table. These are computationally efficient for solving both classification and regression problems [42,43]. However, building a decision table requires careful selection of the most suitable attributes. This is typically carried out by measuring the

table's cross-validation performance for various attribute subsets and selecting the best-performing subset. Currently, to search the attribute space, the "forward" search selection method is used, and "leave one out" cross-validation is employed for evaluating the accuracy of the selected attribute subset (Table 4).

#### 2.4.3. The Random Forest Algorithm

The Random Forest (RF) algorithm is an ensemble non-parametric machine learning algorithm employed to solve classification and regression problems [44]. Bagging and bootstrap sampling are key components of the Random Forest algorithm. Bootstrap sampling helps to introduce randomness and reduce overfitting. Bagging is used to create an ensemble of decision trees, which helps to reduce the variance and improve the accuracy of the model [40]. It works by creating multiple decision trees, where each tree is trained on a different subset of the input data [45]. The trees are created using bootstrap sampling, where each subset is created by randomly selecting data instances from the input data set with replacements. Once the subsets are created, a decision tree is trained on each subset using a random subset of the features. When a new data point is presented to the model, each decision tree makes a prediction, and the final prediction is computed by averaging the predictions of all the decision trees [46,47]. Currently, the Random Tree classifier is used to produce a decision tree that has no depth limit, requires at least one instance in each leaf node and only performs splits that significantly reduce the variance of the target class. The seed for the random number generator is set to 1 for reproducibility, and the capability checking is disabled to speed up the computation. Bagging with 100 iterations is used here (Table 4).

#### 2.4.4. Artificial Neural Network-Multi Layer Perceptron (ANN-MLP)

The Multi-Layer Perceptron (MLP) is a supervised learning algorithm and popular network architecture of an artificial neural network (ANN). It consists of multiple layers of interconnected nodes, with each node performing a simple computation using a weighted sum of its inputs and an activation function [48,49]. The input layer receives the input data, which is then processed through one or more hidden layers before reaching the output layer. MLP is a feed-forward neural network that is trained using the BackPropagation (BP) algorithm. The BP algorithm trains the MLP by adjusting the weights of the connections between the neurons in the network. During training, the algorithm iteratively adjusts the weights to minimize the difference between the predicted output and the actual output, using a method called gradient descent. Moreover, the Sigmoid activation function is commonly used in feed-forward neural networks to introduce non-linearity in the networks typically presented by [50]:

$$y = \frac{1}{1 + \exp(\sum_{j=1}^n w_j x_j)} \quad (2)$$

where  $w_i$  is the weight associated with the  $i$ th input,  $x_j$  is the  $j$ th input value and  $n$  is the number of inputs to the neuron. Hence, MLP is a powerful machine learning algorithm that can be used for both regression and classification tasks. It is capable of learning complex, non-linear relationships between input and output variables, making it suitable for a wide range of applications [51]. Currently, to employ MLP, a dataset will have been pre-processed with the unsupervised attribute normalization filter to scale the attributes and the unsupervised instance removal filter, which randomly removes 20% of the instances (Table 4).

#### 2.4.5. ML Performance

To evaluate the performance of ML algorithms based on the selected scenarios (Table 3), five indicators were employed, as shown in Table 5, and the Taylor diagram was used.

**Table 5.** Selected indicators for the assessment of the performance of the machine learning algorithms.

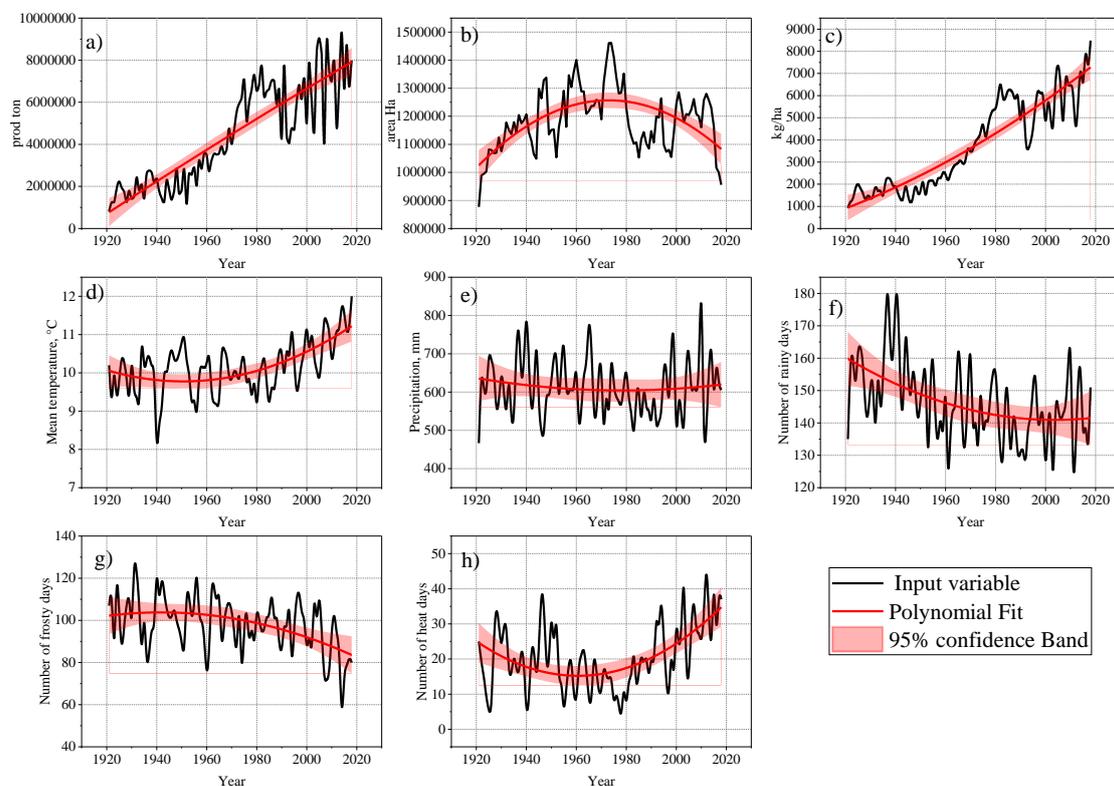
Indicator	Equation *	Range
Correlation coefficient	$r = \left[ \frac{\sum_{i=1}^N \{ (y_A^i - \bar{y}_A)(y_P^i - \bar{y}_P) \}}{\sqrt{\sum_{i=1}^N (y_A^i - \bar{y}_A)^2} \sqrt{\sum_{i=1}^N (y_P^i - \bar{y}_P)^2}} \right]$	[-1 to +1]
Mean absolute error	$MAE = \frac{1}{N} \sum_{i=1}^N  y_P^i - y_A^i $	[0 to ∞]
Root mean squared error	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_P^i - y_A^i)^2}$	[0 to ∞]
Relative absolute error	$RAE = \left  \frac{y_A^i - y_P^i}{y_P^i} \right  \times 100$	[0 to ∞]
Root relative squared error	$RRSE = \frac{\sqrt{\sum_{i=1}^N (y_P^i - y_A^i)^2}}{\sqrt{\sum_{i=1}^N (y_A^i - \bar{y}_A)^2}}$	[0 to ∞]

\*  $y_A^i$ : recorded maize yield,  $y_P^i$ : predicted maize yield,  $\bar{y}$ : mean of recorded maize yield, and  $N$  is the total number of recorded maize yield.

### 3. Results

#### 3.1. Trend Analysis of Variables

Trend analysis of explanatory or predictor variables revealed a rising trend of maize area and production with significant Mann–Kandall (Tau) values of 0.12 and 0.71 ( $p$  value = 0.07, <0.001) and Sen’s slope values of 817.32 and 74,620, respectively, over the period of 98 years. Similar positive rising trends were observed for mean temperature and hot days, with significant ( $p$  value = <0.001, 0.02) Tau values of 0.26 and 0.15, and Sen’s slope values of 0.01 and 0.094, respectively. However, the Sequential Mann–Kandall SQMK test identified 1923 and 1964 as trend breaking, or intersection, years for maize area and production. Similarly, 2007, 2009 and 2011 are identified as intersection years of the prograde temperature series (Table 6). The trend breaking intersection years revealed the occurrence of non-linear characteristics in time-series, which is better presented in the polynomial regression fit in Figure 2.



**Figure 2.** The timeline of the explanatory and response variables: (a) PROD, (b) AREA, (c) yield, (d) Tmean, (e) PRCP, (f) RD, (g) FD and (h) HD.

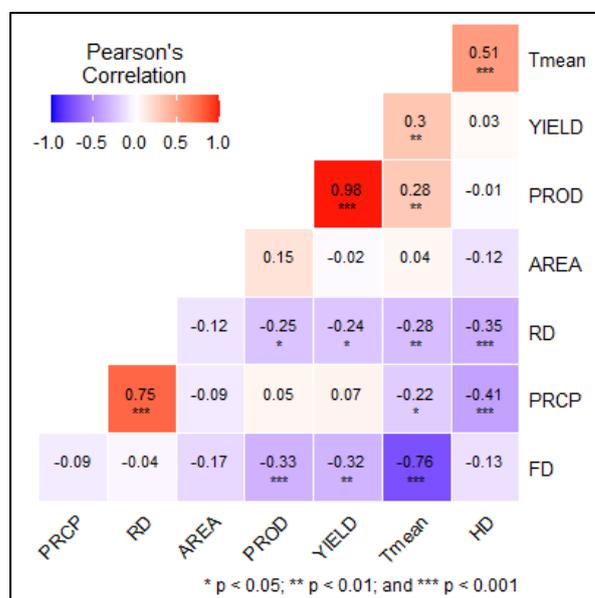
Similarly, precipitation and its associated variables, i.e., rainy and frosty days, followed a declining trend with a negative Sen’s slope of  $-0.24$ ,  $-0.18$  and  $-0.191$  with Tau values of  $-0.041$  ( $p$  value = 0.54),  $-0.233$  ( $p$  value < 0.001) and  $-0.228$  ( $p$  value < 0.001), respectively. As with temperature, several breaking points or trend changing years are observed for precipitation on decadal basis, e.g., 1932–1934 followed by 1946 and 1948. This indicates the years when droughts and floods, or high precipitation, occurred, showing a significant correlation between temperature and precipitation. Another trend in decreasing precipitation is broken in the years 2014–2016 by an above average rainfall of 700 mm in 2016 (Table 6). Hence, the non-linearity of all predictors is more accurately presented in Figure 2 and supports the non-linear ML analysis.

**Table 6.** Trend test (Mann–Kendall, Sen’s slope and SQMK) statistics of all variables.

Variables	Area (hectares)	Production (tons)	Yield (kg/ha)	Mean Temperature (°C)	Precipitation	Rainy Days	Frosty Days	Heat Days
Tau	0.122 (0.07)	0.710 *** (<0.001)	0.726 *** (<0.001)	0.263 *** (<0.001)	-0.041 (0.54)	-0.233 *** (<0.001)	-0.228 *** (<0.001)	0.155 *** (0.02)
z-score	1.78	10.35	10.59	3.77	-0.601	-3.37	-3.31	2.23
Sen’s slope	817.32	74,620	65.48	0.01	-0.24	-0.18	-0.191	0.094
Breaking point	1923	1964	1968	2007, 2009, 2011	1932–1934, 1946, 1948, 1961, 1963, 1967, 2014–2016	1951–1953	2000	2011

Significant \*\*\*  $p < 0.001$ .

Pearson correlation reveals the significance of the linear relationships between climatic and crop-related variables. A significant ( $p < 0.01$ ) positive correlation of 0.3 is observed between the maize yield and Tmean, which explains the increasing mean temperature favoring crop production and yield (Figure 3).



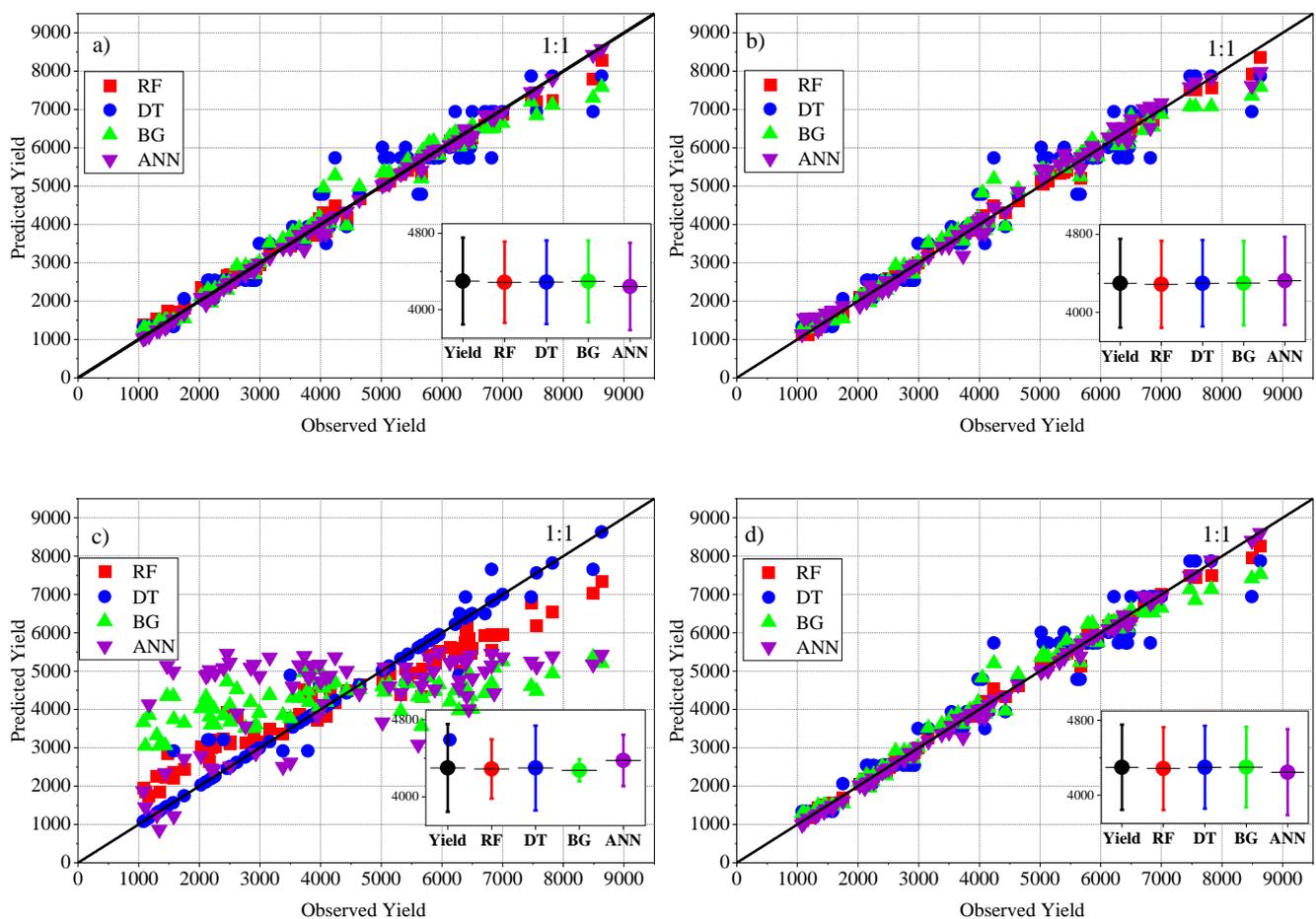
**Figure 3.** Pearson correlation between explanatory and response variables.

Moreover, no significant correlation is observed between precipitation and maize yield, but a negative significant ( $p < 0.05$ ,  $p < 0.01$ ) correlation of 0.24–0.32 is observed between maize yield and rainy and frosty days. A zero or negative correlation between maize area and yield reveals that an increase in maize yield is more associated with other factors such as fertilization or improvement in seed varieties (Figure 3). Hence, the trend and correlation analyses more effectively revealed the positive and negative relationships between all predictors and response variables in the ML analysis. Precipitation does not

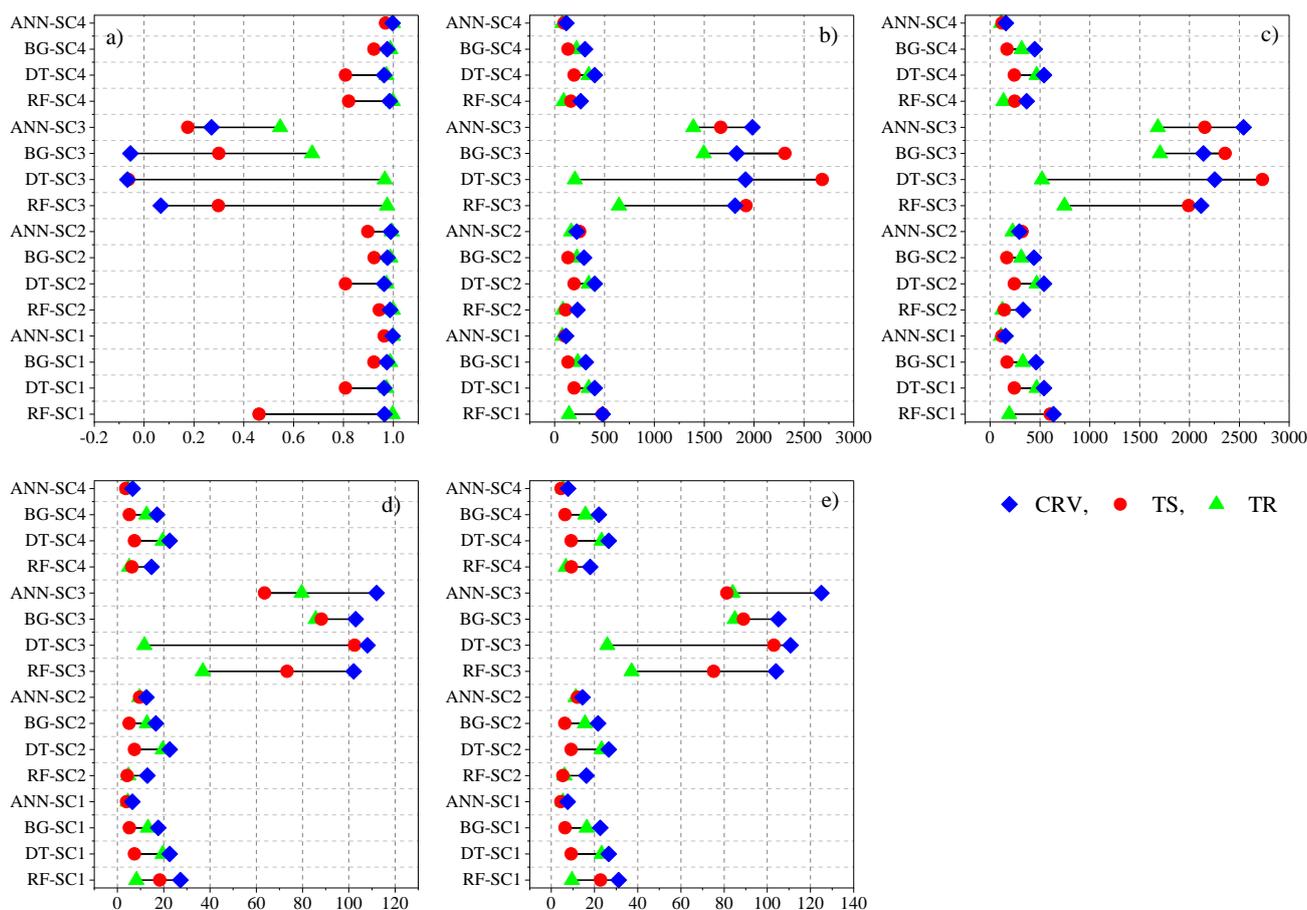
seem to have a significant linear relationship with maize yield; however, in order to explore the hidden, complex and non-linear relationships with maize yield, we did not remove it from the ML analysis.

### 3.2. ML Performance in the Training and Testing Stage

The performance evaluation metrics (Table 4) of all machine learning algorithms in the training stage revealed that ANN-MLP outperformed RF, BG and DT with the highest correlation of  $r = 0.999$  and  $0.998$  in SC1 and SC4, respectively. The lowest RMSE and RAE were 107.9 and 4.53%, respectively, for ANN-MLP-SC1, followed by 110 and 4.46%, respectively, in ANN-MLP-SC4. The ANN-MLP was also the optimum model machine learning algorithm for maize yield prediction in Hungary. Along with MLP, RF also performed competitively for maize yield prediction, with a high correlation of  $r = 0.998$  in both the RF-SC4 and RF-SC2, followed by 0.997 and 0.975 in the RF-SC1 and RF-SC3, respectively. Other evaluation metrics also revealed the lowest RMSE and RAE values of 122.1 and 4.7%, respectively, for RF-SC2; 135 and 5.1%, respectively, for RF-SC4; 192 and 8.2%, respectively, for RF-SC1 and 745.3 and 36.9%, respectively, for RF-SC3 (Figures 4 and 5).



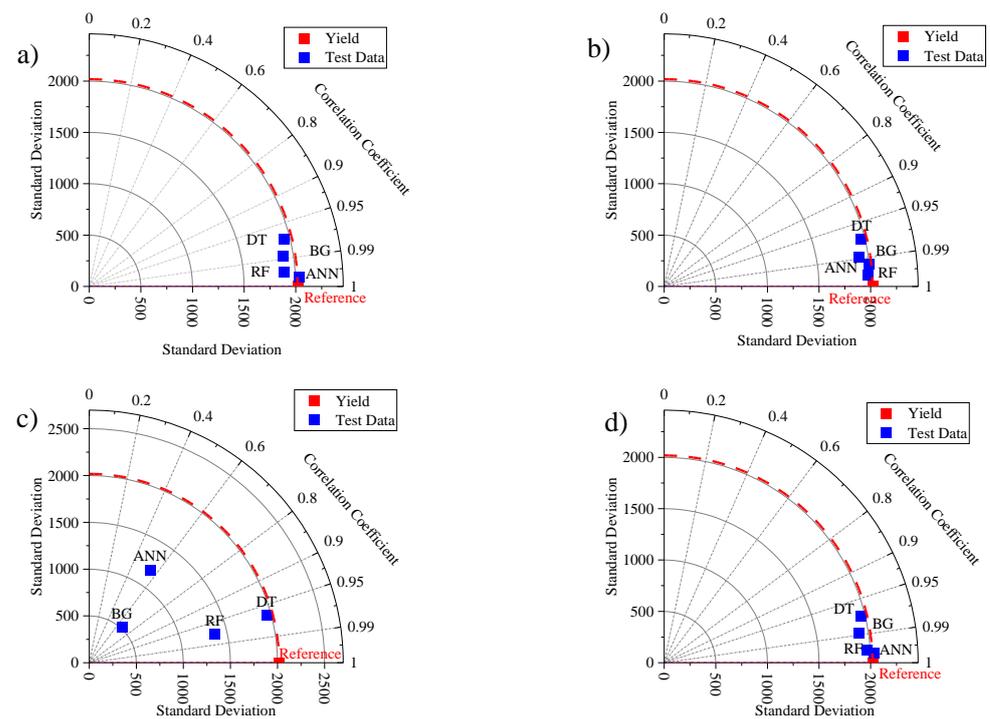
**Figure 4.** A scatter plot between the predicted and the observed yield, based on different scenarios using the four algorithms (Bagging (BG), Decision Table (DT), Random Forest (RF) and Artificial Neural Network-Multi Layer Perceptron (ANN-MLP)): (a) SC1: AREA+ PROD+ Tmean+ PRCP+ RD+ FD+ HD, (b) SC2: AREA+ PROD, (c) SC3: Tmean+ PRCP+ RD+ FD+ HD and (d) SC4: AREA+ PROD+ Tmean+ PRCP.



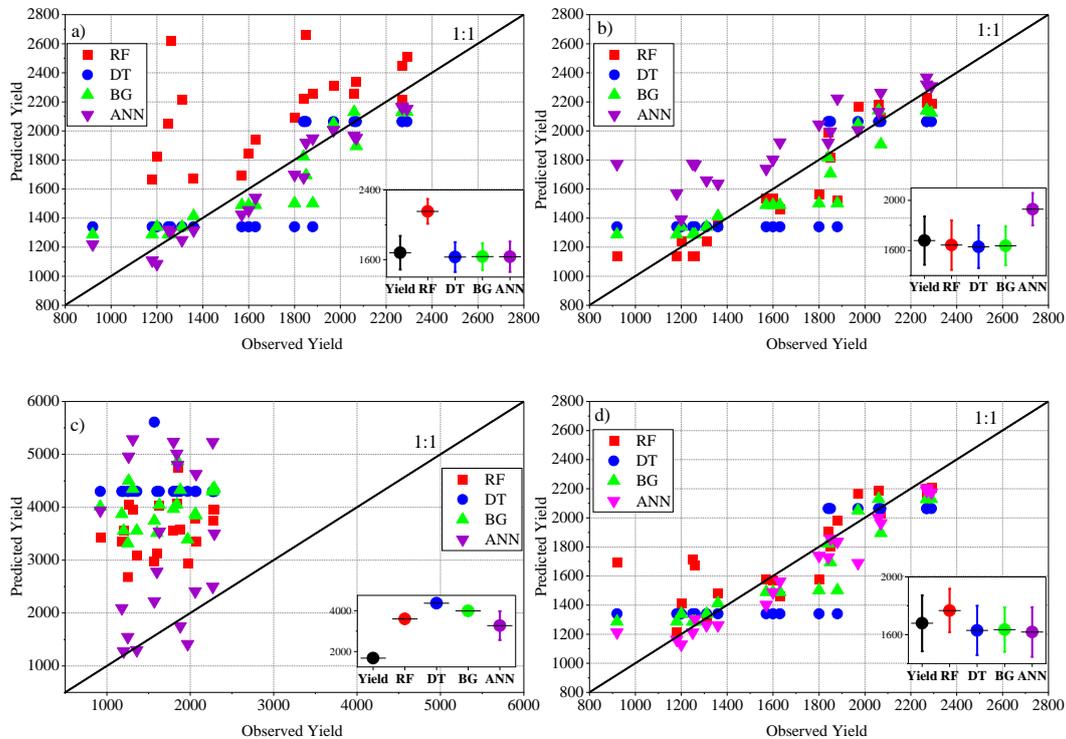
**Figure 5.** The statistical performance of ML under different scenarios in training (TR), testing (TS) and cross validation (CRV): (a) correlation coefficient, (b) mean absolute error, (c) root mean squared error, (d) relative absolute error and (e) root relative squared error.

The Taylor diagram (Figure 6) also revealed that the optimum model for maize prediction was ANN-MLP followed by RF, BG and DT in different scenarios. The DT proved to be the least applicable model in the training stage of yield prediction, with a high RMSE of 467.6 in DT-SC1, DT-SC2 and DT-SC4 and 520.2 in DT-SC3. Hence, at the training stage, the optimum ML algorithm with the highest accuracy of yield prediction was ANN-MLP for the optimum scenarios of SC1 and SC4 (Figures 4 and 5), which are further evaluated at the testing and cross validation stages.

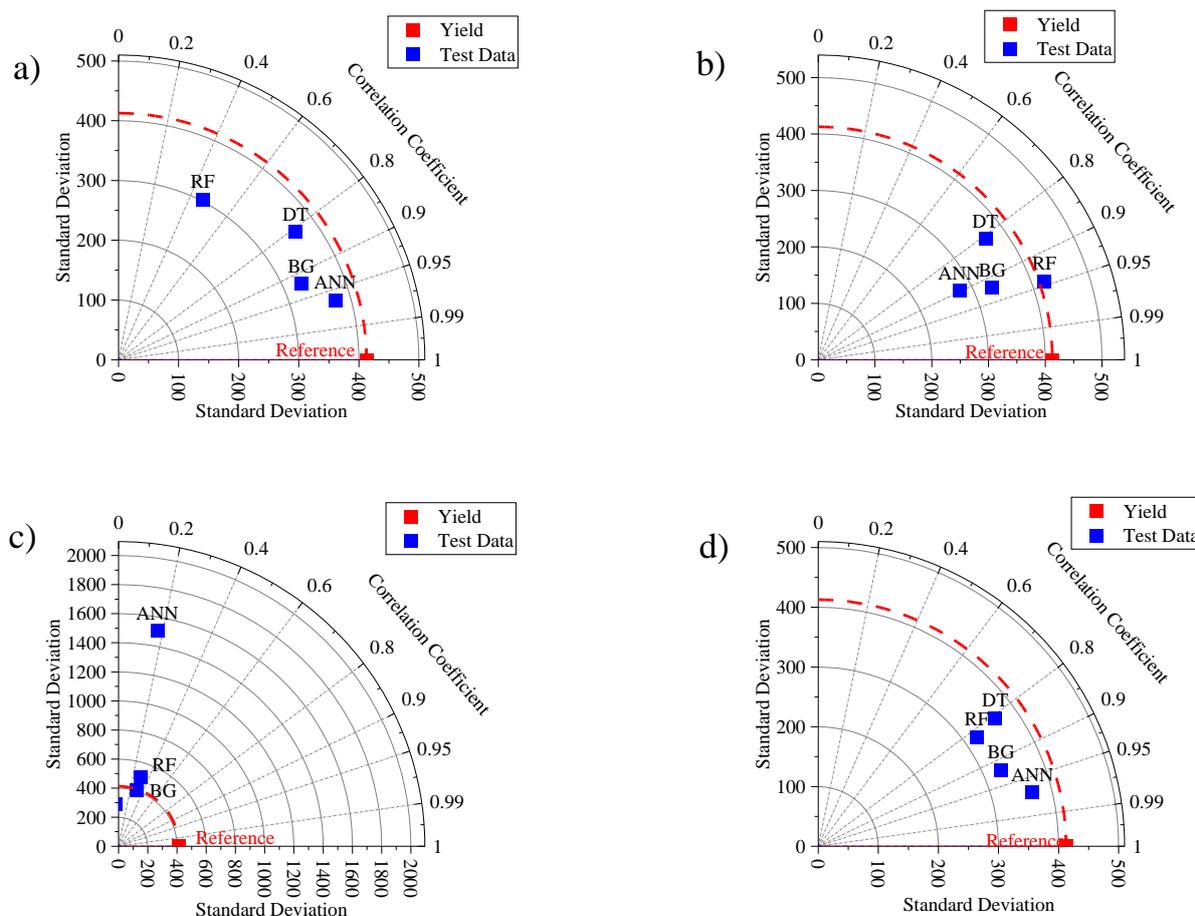
Following the training stage, the performance evaluation metrics at the testing stage also revealed the outperformance of the ANN-MLP algorithm over other ML algorithms, and it was selected as an optimum model for yield prediction at the regional level. The highest correlation  $r = 0.96$ , RMSE = 120 and RAE = 3.6% was recorded for ANN-MLP-SC4, which was followed by  $r = 0.94$ , RMSE = 118.9 and RAE = 3.98% for ANN-MLP-SC1. The relative values for RF-SC2 were  $r = 0.94$ , RMSE = 141 and RAE = 4.2%, followed by RF-SC4, where  $r = 0.82$ , RMSE = 246.8 and RAE = 6.27% (Figure 5). Furthermore, SC3 was found to perform poorly at the testing stage for optimum maize yield prediction (Figure 7), with the lowest correlation and highest RMSE in all the ML algorithms. BG also performed with a good accuracy of  $r = 0.92$ , and an RMSE = 167.4 and 168.9 for SC2 and SC4, respectively, and SC1 with the least  $r = 0.3$ , and the highest RMSE and RAE = 2359 and 88%, respectively, for SC3. The Taylor diagram in the testing stage (Figure 8) also revealed the outperformance of ANN-MLP over the other ML algorithms, with the highest correlation and accurate prediction for SC4 and SC1, followed by BG for SC4, SC1 and SC2 and RF for SC2 and SC4. Hence, the overall performance of the ML algorithms for yield prediction is sequenced as ANN-MLP > BG > RF > DT and the scenarios are sequenced as SC4 > SC1 > SC2 > SC3.



**Figure 6.** A Taylor diagram showing the performance of ML in the training stage under different scenarios: (a) SC1: AREA+ PROD+ Tmean+ PRCP+ RD+ FD+ HD, (b) SC2: AREA+ PROD, (c) SC3: Tmean+ PRCP+ RD+ FD+ HD and (d) SC4: AREA+ PROD+ Tmean+ PRCP.



**Figure 7.** A scatter plot of the testing phase between the predicted and observed yield as based on different scenarios using the four algorithms (Bagging (BG), Decision Table (DT), Random Forest (RF) and Artificial Neural Network-Multi Layer Perceptron (ANN-MLP)): (a) SC1: AREA+ PROD+ Tmean+ PRCP+ RD+ FD+ HD, (b) SC2: AREA+ PROD, (c) SC3: Tmean+ PRCP+ RD+ FD+ HD and (d) SC4: AREA+ PROD+ Tmean+ PRCP.



**Figure 8.** A Taylor diagram for the testing phase showing the performance of ML under different scenarios: (a) SC1: AREA+ PROD+ Tmean+ PRCP+ RD+ FD+ HD, (b) SC2: AREA+ PROD, (c) SC3: Tmean+ PRCP+ RD+ FD+ HD and (d) SC4: AREA+ PROD+ Tmean+ PRCP.

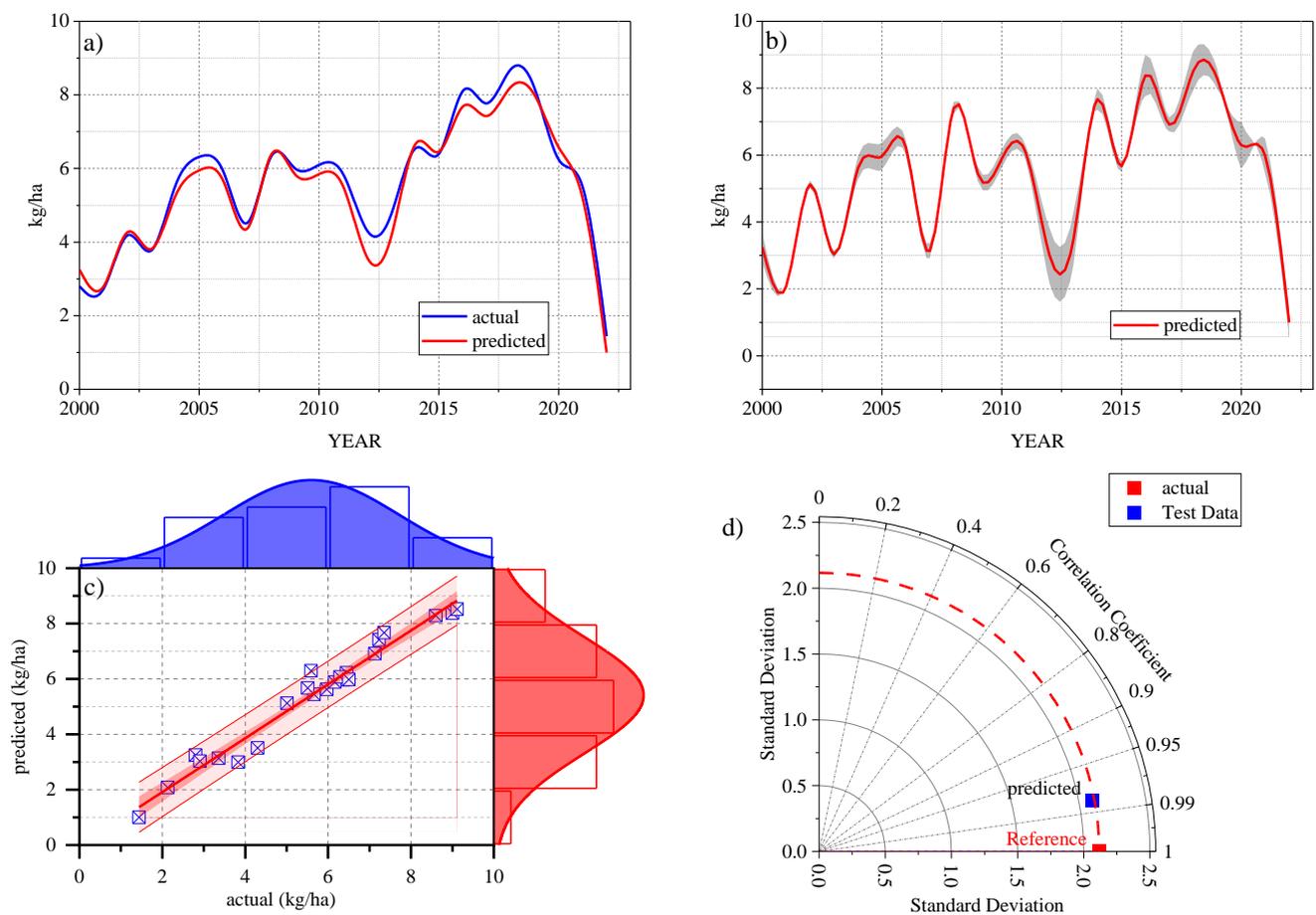
### 3.3. ML Performance in the Cross-Validation Stage

Cross validation revealed strong competition between all ML algorithms for accurate yield prediction. However, the highest correlation of  $r = 0.997$ , with a low RMSE and RAE = 158 and 6.5%, respectively, was for the ANN-MLP-SC4, followed by the ANN-MLP-SC1 with  $r = 0.997$  and a low RMSE and RAE = 153 and 6.4%, respectively, and ANN-MLP-SC2 with  $r = 0.990$  and RMSE and RAE = 293 and 12.4%, respectively (Figure 5). Subsequently, RF outperformed with the highest  $r = 0.987$ , RMSE = 330.3 and RAE = 12.9% for RF-SC2, followed by  $r = 0.985$ , RMSE = 366.4 and RAE = 14.7% for RF-SC4 and  $r = 0.965$ , RMSE = 634.4 and RAE = 27.1% for RF-SC1. BG performed with the highest  $r = 0.976$ , RMSE = 440 and RAE = 16.6% for BG-SC2, followed by  $r = 0.975$ , RMSE = 447 and RAE = 17% for BG-SC4. Overall, the model performance in accurate yield prediction from cross-validation is sequenced as ANN-MLP > RF > BG > DT (Figure 5).

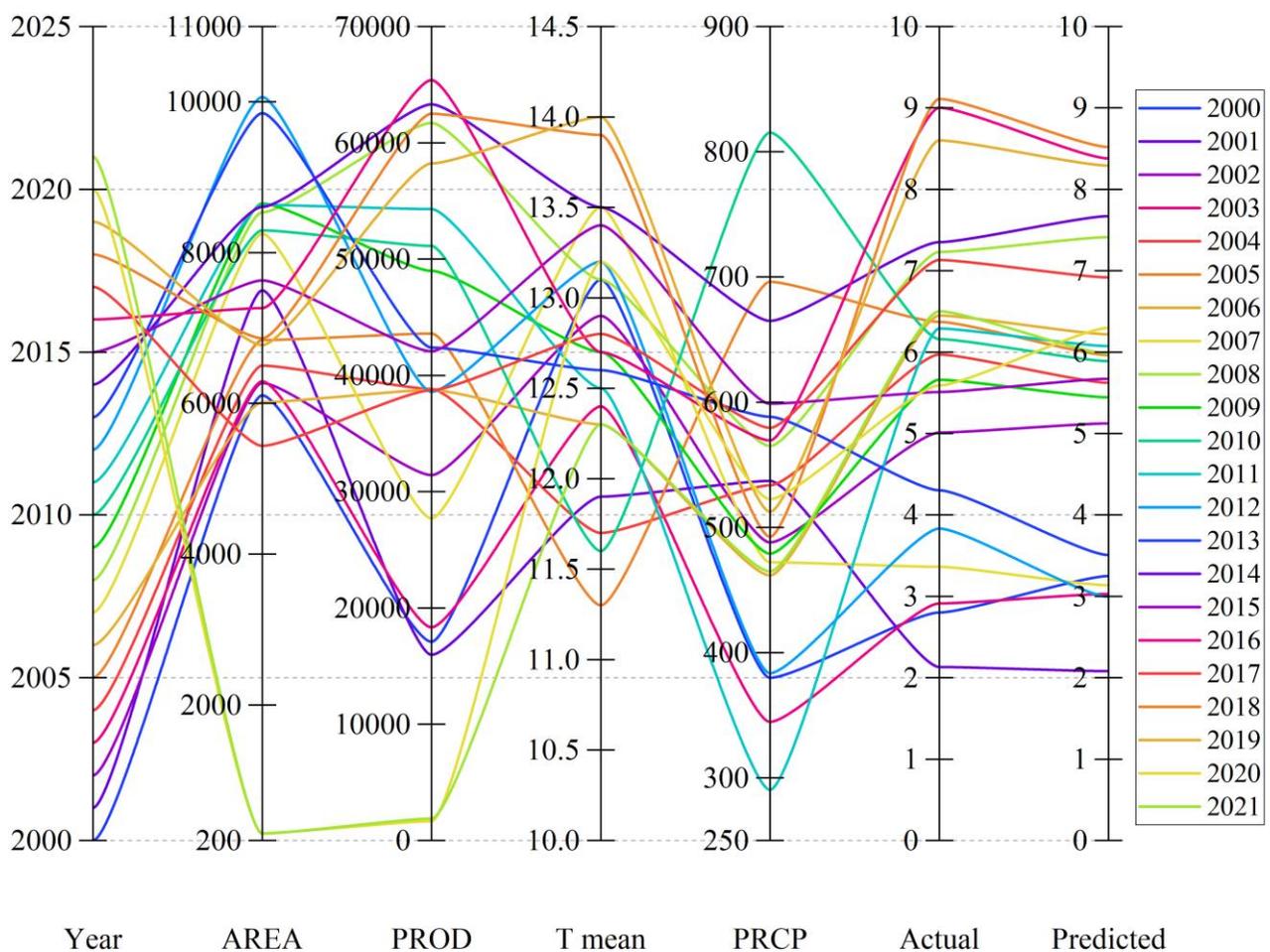
Application of all the ML algorithms revealed that ANN-MLP outperformed the other algorithms, proving to be the optimum model for maize yield prediction. Among the four tested scenarios, SC4 and SC1 were found to be more accurate for maize yield prediction in training, testing and cross-validation. Hence, considering the competition between all climatic and agricultural factors in the different scenarios, the explanatory factors of SC4 (AREA, PROD, Tmean and PRCP) are the most significant predictors for maize yield in Hungary.

### 3.4. Exploring the Flexibility of Machine Learning Algorithms for Predicting Maize Yield at a Regional Scale in Hungary

The ANN-MLP-SC<sub>4</sub> was the most powerful optimum algorithm/scenario among the other tested algorithms/scenarios in Table 2. It was used to validate its own performance in predicting maize yield at a regional scale, where Budapest County was chosen as a test region. The ANN-MLP-SC<sub>4</sub> reached a high-performance standard ( $r = 0.98$ , relative absolute error = 21.87%, root relative squared error = 20.45% and RMSE = 423.23) as shown in Figure 9. The Taylor diagram exhibited an excellent performance of ANN-MLP-SC<sub>4</sub> for predicting maize yield (Figure 9d). In this sense, Figure 10 shows the timeline changes among different input variables for SC<sub>4</sub> from 2000 to 2021, where each year is assigned to a different color. Through this we can track the changes in the variables and evaluate the margin between the observed and predicted values. In this context, the relationship between the actual and predicted maize yields are similar in most years. However, instances of over prediction were identified in 2008 and 2014.



**Figure 9.** The outlook of ANN-MLP-SC<sub>4</sub> performance in predicting maize yield at Budapest County: (a) timeline of actual and predicted values, (b) errors in predicting maize yield (gray shadow), (c) scatter plot between the actual and the predicted maize yield and (d) the Taylor diagram.



**Figure 10.** An overview of SC4 input variables (AREA+ PROD+ Tmean+ PRCP) and ANN-MLP-SC4 output.

#### 4. Discussion

##### 4.1. Climatic Influence on Maize Yield Prediction

Climatic changes characterized by extreme events such as droughts and heat stress are projected to intensify. These affect the growth cycles and development of the many cereal crops such as maize [52]. Climate variability is reported to explain over 60% of crop yield variability in the substantial 'global breadbasket' areas [53].

Our study explored the climatic factors in different scenarios for yield prediction modelling at a regional scale in Central Europe. Among five climatic input variables, i.e., Tmean, PRCP, HD, RD and FD, the Tmean was found to have the most significant influence on maize yield, with a positive correlation, and was also found to be a good predictor, in tandem with PRCP, for yield prediction modelling in SC4 (Figures 3, 4 and 7). For instance, Shahhosseini, Martinez-Feria, Hu and Archontoulis [18] employed ML methods and also identified temperature and rainfall as significant predictors of maize yield. Another study conducted by Meng, et al. [54] used random forest regression and reported maximum temperature and rainfall as relatively important climatic predictors of maize yield at the farm level. The significant ( $p < 0.001$ ) rising trend of Tmean and maize yield (Table 3), as well as their positive correlation (Figure 3), affirms the findings that the increase in temperature favors the growth of maize in Europe and other regions [55]. Overall, maize requires a moderately high temperature to grow, but an optimal temperature above 35 °C increases the rate of transpiration and causes yield loss [56]. Furthermore, another study by Bussay, van der Velde, Fumagalli and Seguni [32] also showed that the favorable agroclimatic conditions of Hungary have made maize the most widely grown crop at the national level. However, temperature-related climatic factors may not always favor the growth of this

crop, e.g., ref. [57] reported that climatically derived growing degree days (GDD) were negatively correlated with maize yield in China. Furthermore, Hatfield [58] noted that higher night temperatures increased the senescence rate, which shortened the grain-filling period. Accordingly, kernel number reduction occurs if plants are exposed to over 30 °C during the pollination stage. Similarly, elevated temperatures cause block tassels and ears, elongate the anthesis and silking interval, damage anther structure, reduce pollen activity and grain number and consequently reduce yield [56]. Hence, our findings clearly reveal that mean temperature is the significant predictor of maize yield at the country level, which can be explored further by combining the minimum and maximum temperature input at different stages of maize growth. Among the four tested scenarios, the variables of SC1 were found to be the second-most significant for maize yield prediction (Figure 8). Hence, other than temperature, HD also explains the variability of maize yield in a non-linear relationship employing machine learning methods. In terms of heat stress, Edreira, et al. [59] noted that the heat effect on the kernel weight of temperate maize hybrids occurred in the first half of the effective grain filling period, rather than around flowering, by improving the availability of assimilates to each kernel and to the carbohydrate reserves in the stem at physiological maturity. An investigation by Lizaso, et al. [60] revealed that heat stress ranging from 42.9–52.5 °C did not extend the anthesis–silking interval, though it shortened the vegetative and reproductive phases (with a whole cycle of 30 days). A reduction in yield mainly occurred due to the reduced pollen viability.

Other than the temperature and hot days, the precipitation characteristics, such as the number of rainy days and the annual amount (mm) and timing of rain, also affect the maize yield and were found to be significant predictors in the different scenarios (Figure 3). The decline in rainfall shortens the growing season, which limits water and nutrients, causing a decline in yield of 15 to 20% [61], especially in rain-fed systems. Omoyo, et al. [62] explains that variability in the time to the onset and cessation of rainfall influences the maize growth cycle. The findings in our study suggest no linear correlation between PRCP and maize yield (Figure 3); however, the ML identified that in scenarios SC4 and SC1, PRCP was a significant non-linear predictor of maize yield (Figure 8). This concurs with other studies on ML-based maize yield prediction [18,54]. The impacts of PRCP, or rain, in predicting the maize yield can be discussed in either way. A strong positive correlation between the quantity of rainfall with maize yield was reported in Nigeria, with the rainfall characteristics conjointly accounting for 67.4% of the yield [63]. Additionally, Kern, Barcza, Marjanović, Árendás, Fodor, Bónis, Bognár and Lichtenberger [31] reported the positive correlation between PRCP and maize yield in Hungary, in that there was sufficient soil moisture for improving or enhancing the yield.

However, the Mann–Kendall trend analysis of climatic predictors in our study reported a slight decline in precipitation (Table 3), which affirmed the occurrence of intense, frequent drought events affecting the maize and wheat yields in Hungary in recent decades [30]. Despite frequent droughts in recent years, improvements in maize yields are associated with several non-climatic factors, including planned land use adaptation, appropriate irrigation scheduling, improved fertilization and the adaptation of precision farming methods [64–66]. Frost is another climate element that affects plant structure and reproduction. Severe frosts cause the loss of leaves and disrupt active physiological processes, such as the translocation of assimilates to ears and the conversion of sugars to starch in the kernels, hence inhibiting kernel filling [67]. Overall, the dominance of the effect of climate change on maize growth and yield is not unidirectional. For example, in the U.S., the effect of extreme temperatures on controlling the soil water demand and transpiration rate in rainfed maize made temperature a more effective predictor of grain yield [53,68]. There is clearly the need for further investigation into the prediction of maize growth in relation to the growth cycle and yield response to climate change.

#### 4.2. Machine Learning Algorithms for Better Optimization of Maize Yield

Crop yield prediction is a critical task for farmers and policymakers, supporting them to make informed decisions on planting, harvesting and food security. In recent years, data mining techniques and machine learning (ML) algorithms have shown great promise in predicting maize yield. Accurate yield prediction employing various machine and deep learning algorithms is widely reported by researchers [69–74]. In this context, we tested the performance of four ML algorithms including BG, DT, RF and ANN-MLP (Figures 4–7). The random forest algorithm is often preferentially applied in data mining and yield prediction studies due to its robustness to noisy data, high accuracy and its utilization of the bagging technique [75,76]. Bagging creates an ensemble of decision trees, which reduces the variance and provides a more accurate predicted model [27]. The results of our study proposed a high performance of RF over BG and DT, with a high correlation and low RMSE in the training and cross validation stages (Figure 5). Keerthana, et al. [77] also reported that the performance of ensemble algorithms was superior for crop yield prediction. Similar high performance results of RF over linear regressions are reported by Jeong, et al. [78] for the accurate crop yield predictions of several crops (wheat, maize and potato) from climatic and biophysical variables at regional and global scales. Another study [25] demonstrated RF to be an optimized algorithm for maize yield prediction at the county level, combining multisource variables. However, several hyperparameters of ANN-MLP, e.g., the number of hidden layers, number of neurons per layer, the activation function and learning rate, make it more capable and flexible in exploring the hidden non-linear relationships between variables [79]. A recent study by Ahmed [80] proposed a modified MLP with Spider Monkey Optimization, which provided a promising performance in maize yield prediction. The results of our study report ANN-MLP as the optimum algorithm with a high accuracy of yield prediction in SC4, followed by SC1 (Figure 5). However, a strong competition existed between the performance of RF and ANN-MLP, with a slight difference in the correlation for SC4 and SC1 (Figures 6 and 8), proving them to be optimal for maize yield prediction. ANN-MLP has proven superior for exploring the non-linear relationships between variables and capturing the complex relationships between predictors and response variables, which otherwise needed classical statistical approaches to be understood [81]. The outperformance of ANN-MLP over other ML algorithms in our study also concurs with other studies reporting accurate yield predictions from ANN over other ML methods [82,83]. For instance, Bhojani and Bhatt [84], in their study of wheat yield prediction, described MLP as the optimum algorithm, with new modifications in its activation functions in the WEKA environment. Another study by Hara, et al. [85] also preferred to use ANNs for crop yield prediction from climatic and remote sensing data. Hence, the machine and deep learning algorithms proved to be efficient techniques for yield forecasting at a regional scale in Europe.

#### 5. Research Limitations and Model Uncertainties

Overall, machine learning provides a more robust and optimistic approach for accurate yield prediction and other agriculture management studies. Our study also provided the meaningful application of ML algorithms in a Central European region with fewer similar studies. The results of our study provided a baseline for local stakeholders and policy makers to ensure better mitigation and adaptation strategies to cope with the negative impacts of future climatic change. ANN-MLP was proven to be the optimum model for accurate yield prediction at a regional level, followed by RF. Uncertainties in model implementation and explanatory variables created certain limitations. Our study evaluated maize yield forecasting in a sample of 98 observations from a historical time-series between 1921 and 2018. Overall, the 10-fold cross validation approach gave robust and accurate predictions, producing reliable results. Previous studies have also forecast crop yield prediction from a historical time-series [86–88]; however, including time-series observations at the county or state level can enhance the sample size and improve the performance accuracy [88]. Furthermore, we employed the predictors in different scenarios (SC1 to

SC4) to examine their significance in yield prediction and identified temperature and precipitation as the most significant climatic predictors in combination with the agricultural inputs. However, for a more detailed analysis of the predictors, a sensitivity analysis of relative importance can be conducted by increasing the number of predictors or the explanatory variables. For example, the utilization of several other climatic variables such as minimum temperature, maximum temperature, vapor pressure and the active and short-wave radiations can provide more detailed explanations for maize yield prediction, as well as sensitivity analyses, partial correlations or measuring the relative importance [89]. The further addition of non-climatic factors such as fertilization, irrigation scheduling and land management practices could also be incorporated in future research to better explain the yield variability on a specific time scale. Moreover, several remote sensing variables such as the normalized difference vegetation index (NDVI) could also be considered in combination with climatic response variables for better yield estimation [90]. In the context of ML applicability, model uncertainties can be improved by feature selection in big data, hyperparameter tuning, regularization and the ensemble model approach (combining multiple ML algorithms) to achieve accuracy in yield prediction.

## 6. Conclusions

Machine learning (ML) algorithms are a flexible tool for predicting crop yield by compiling specific data related to crop characteristics, agricultural practices, climate and soil. Compared with classical crop modeling programs, the ML algorithms can identify patterns and make predictions more easily and accurately, thus enhancing the crop management system. In this research, four ML algorithms (BG, DT, RF and ANN-MLP) and four scenarios were employed to select the optimum combination to facilitate accurate decision making. The output of the research is summarized as follows:

1. Maize production and area were significantly increased across Hungary between 1921 and 2018, as shown by Sen's slopes of 1.78 and 10.35 ( $p < 0.05$ ), respectively. Similar positive rising trends were observed for the Tmean and HD (Sen's slopes: of 0.01 and 0.094, respectively). In contrast, PRCP, RD and FD exhibited declining trends with negative Sen's slopes of  $-0.24$ ,  $-0.18$  and  $-0.191$ , respectively.
2. In the training stage, the majority of the algorithms showed a high flexibility in predicting maize yield regardless of the applied scenario, where the  $r$  value ranged between 0.99 (ANN-MLP-SC4, ANN-MLP-SC1, RF-SC1, RF-SC2, RF-SC3 and RF-SC4) and 0.54 (ANN-MLP-SC3), and the RMSE ranged between 107.9 (ANN-MLP-SC1) and 1704.8 (BG-SC3);
3. In the testing stage, the performance of the ML in predicting maize yield varied significantly. The highest performance was recorded in ANN-MLP-SC4 ( $r = 0.96$ , RMSE = 120 and RAE = 3.6%), followed by ANN-MLP-SC1 ( $r = 0.94$ , RMSE = 118.9 and RAE = 3.98%) and RF-SC2 ( $r = 0.94$ , RMSE = 141 and RAE = 4.2%);
4. Based on the 10-fold cross validation, the ANN-MLP-SC4 showed a high ability to predict maize yield with the values  $r = 0.99$ , RMSE = 158 and RAE = 6.5%.
5. The implementation of ANN-MLP-SC4 at a regional scale (Budapest region) was highly successful with the values  $r = 0.98$ , RMSE = 423 and RAE = 21.5%.

Overall, the output of this research recommends ANN-MLP as an optimum model for maize yield prediction at a regional level combining climate (Tmean, PRCP) and agricultural/crop data (AREA, PROD). Hence, this research identifies an effective tool for developing sustainable agricultural management strategies against climate change. Furthermore, precision agriculture technology could benefit from this research by adopting ANN-MLP for field-level studies. This study also contributed to predicting wheat yield using ML algorithms based on different data sources, such as remote sensing data, climate data and crop management data.

**Author Contributions:** Conceptualization, E.H. and S.M.; methodology, S.M.; validation, S.A. and S.M.; formal analysis, O.H. and S.M.; investigation, A.S.; resources, A.V.; data curation, S.M.; writing—original draft, S.A. and A.O.; writing—review and editing, B.B., A.V., A.A. and I.B.; visualization, T.R.; funding acquisition, E.H. All authors have read and agreed to the published version of the manuscript.

**Funding:** Project no. TKP2021-NKTA-32 has been implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development, and Innovation Fund, financed under the TKP2021-NKTA funding scheme. Additionally, this research was supported by the Researchers Supporting Project number (RSP2023R296), King Saud University, Riyadh, Saudi Arabia.

**Data Availability Statement:** Data available at: <https://www.ksh.hu/> (accessed on 1 April 2021).

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

## References

1. Raheem, D.; Dayoub, M.; Birech, R.; Nakiyemba, A. The Contribution of Cereal Grains to Food Security and Sustainability in Africa: Potential Application of UAV in Ghana, Nigeria, Uganda, and Namibia. *Urban Sci.* **2021**, *5*, 8. [CrossRef]
2. FAO. *The Future of Food and Agriculture: Trends and Challenges*; FAO: Rome, Italy, 2017; pp. 1–163.
3. Erenstein, O.; Jaleta, M.; Sonder, K.; Mottaleb, K.; Prasanna, B.M. Global maize production, consumption and trade: Trends and R&D implications. *Food Secur.* **2022**, *14*, 1295–1319. [CrossRef]
4. Shiferaw, B.; Prasanna, B.M.; Hellin, J.; Bänziger, M. Crops that feed the world 6. Past successes and future challenges to the role played by maize in global food security. *Food Secur.* **2011**, *3*, 307–327. [CrossRef]
5. Grote, U.; Fasse, A.; Nguyen, T.T.; Erenstein, O. Food Security and the Dynamics of Wheat and Maize Value Chains in Africa and Asia. *Front. Sustain. Food Syst.* **2021**, *4*, 617009. [CrossRef]
6. Ray, D.K.; Mueller, N.D.; West, P.C.; Foley, J.A. Yield Trends Are Insufficient to Double Global Crop Production by 2050. *PLoS ONE* **2013**, *8*, e66428. [CrossRef]
7. Okolie, C.C.; Danso-Abbeam, G.; Groupson-Paul, O.; Ogundeji, A.A. Climate-Smart Agriculture Amidst Climate Change to Enhance Agricultural Production: A Bibliometric Analysis. *Land* **2023**, *12*, 50. [CrossRef]
8. Nelson, G.C.; Rosegrant, M.W.; Koo, J.; Robertson, R.; Sulser, T.; Zhu, T.; Ringle, C.; Msangi, S.; Palazzo, A.; Batka, M. *Climate Change: Impact on Agriculture and Costs of Adaptation*; International Food Policy Research Institute: Washington, DC, USA, 2009; Volume 21.
9. FAO. *FAOSTAT Crop Database*; FAO: Rome, Italy, 2019.
10. Prasanna, B.M.; Cairns, J.E.; Zaidi, P.H.; Beyene, Y.; Makumbi, D.; Gowda, M.; Magorokosho, C.; Zaman-Allah, M.; Olsen, M.; Das, A.; et al. Beat the stress: Breeding for climate resilience in maize for the tropical rainfed environments. *Theor. Appl. Genet.* **2021**, *134*, 1729–1752. [CrossRef]
11. Tigchelaar, M.; Battisti, D.S.; Naylor, R.L.; Ray, D.K. Future warming increases probability of globally synchronized maize production shocks. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6644–6649. [CrossRef]
12. Senapati, N.; Halford, N.G.; Semenov, M.A. Vulnerability of European wheat to extreme heat and drought around flowering under future climate. *Environ. Res. Lett.* **2021**, *16*, 024052. [CrossRef]
13. Trnka, M.; Rötter, R.P.; Ruiz-Ramos, M.; Kersebaum, K.C.; Olesen, J.E.; Žalud, Z.; Semenov, M.A. Adverse weather conditions for European wheat production will become more frequent with climate change. *Nat. Clim. Chang.* **2014**, *4*, 637–643. [CrossRef]
14. Webber, H.; Ewert, F.; Olesen, J.E.; Müller, C.; Fronzek, S.; Ruane, A.C.; Bourgault, M.; Martre, P.; Ababaei, B.; Bindi, M.; et al. Diverging importance of drought stress for maize and winter wheat in Europe. *Nat. Commun.* **2018**, *9*, 4249. [CrossRef] [PubMed]
15. Pant, J.; Pant, R.P.; Kumar Singh, M.; Pratap Singh, D.; Pant, H. Analysis of agricultural crop yield prediction using statistical techniques of machine learning. *Mater. Today: Proc.* **2021**, *46*, 10922–10926. [CrossRef]
16. van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. *Comput. Electron. Agric.* **2020**, *177*, 105709. [CrossRef]
17. Shahhosseini, M.; Hu, G.; Archontoulis, S.V. Forecasting Corn Yield With Machine Learning Ensembles. *Front. Plant Sci.* **2020**, *11*, 1120. [CrossRef]
18. Shahhosseini, M.; Martinez-Feria, R.A.; Hu, G.; Archontoulis, S.V. Maize yield and nitrate loss prediction with machine learning algorithms. *Environ. Res. Lett.* **2019**, *14*, 124026. [CrossRef]
19. Bazrafshan, O.; Ehteram, M.; Moshizi, Z.G.; Jamshidi, S. Evaluation and uncertainty assessment of wheat yield prediction by multilayer perceptron model with bayesian and copula bayesian approaches. *Agric. Water Manag.* **2022**, *273*, 107881. [CrossRef]
20. Cao, J.; Zhang, Z.; Luo, Y.; Zhang, L.; Zhang, J.; Li, Z.; Tao, F. Wheat yield predictions at a county and field scale with deep learning, machine learning, and google earth engine. *Eur. J. Agron.* **2021**, *123*, 126204. [CrossRef]
21. Cubillas, J.J.; Ramos, M.I.; Jurado, J.M.; Feito, F.R. A Machine Learning Model for Early Prediction of Crop Yield, Nested in a Web Application in the Cloud: A Case Study in an Olive Grove in Southern Spain. *Agriculture* **2022**, *12*, 1345. [CrossRef]
22. Kamath, P.; Patil, P.; Shrilatha, S.; Sushma; Sowmya, S. Crop yield forecasting using data mining. *Glob. Transit. Proc.* **2021**, *2*, 402–407. [CrossRef]

23. Bolker, B.M.; Brooks, M.E.; Clark, C.J.; Geange, S.W.; Poulsen, J.R.; Stevens, M.H.H.; White, J.-S.S. Generalized linear mixed models: A practical guide for ecology and evolution. *Trends Ecol. Evol.* **2009**, *24*, 127–135. [[CrossRef](#)] [[PubMed](#)]
24. Otukei, J.R.; Blaschke, T. Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *Int. J. Appl. Earth Obs. Geoinf.* **2010**, *12*, S27–S31. [[CrossRef](#)]
25. Pham, H.; Olafsson, S. Bagged ensembles with tunable parameters. *Comput. Intell.* **2019**, *35*, 184–203. [[CrossRef](#)]
26. Mohammed, S.; Alsafadi, K.; Enaruvbe, G.O.; Bashir, B.; Elbeltagi, A.; Széles, A.; Alsalman, A.; Harsanyi, E. Assessing the impacts of agricultural drought (SPI/SPEI) on maize and wheat yields across Hungary. *Sci. Rep.* **2022**, *12*, 8838. [[CrossRef](#)] [[PubMed](#)]
27. 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](#)]
28. Bussay, A.; van der Velde, M.; Fumagalli, D.; Seguíni, L. Improving operational maize yield forecasting in Hungary. *Agric. Syst.* **2015**, *141*, 94–106. [[CrossRef](#)]
29. Pinke, Z.; Lövei, G.L. Increasing temperature cuts back crop yields in Hungary over the last 90 years. *Glob. Chang. Biol.* **2017**, *23*, 5426–5435. [[CrossRef](#)]
30. Kendall, M.G. *Rank Correlation Methods*; Charles Griffin: London, UK, 1948.
31. Mann, H.B. Nonparametric tests against trend. *Econom. J. Econom. Soc.* **1945**, *3*, 245–259. [[CrossRef](#)]
32. Sneyers, R. *On the Statistical Analysis of Series of Observations*; World Meteorological Organization: Geneva, Switzerland, 1991.
33. Cao, J.; Zhang, Z.; Tao, F.; Zhang, L.; Luo, Y.; Han, J.; Li, Z. Identifying the contributions of multi-source data for winter wheat yield prediction in China. *Remote Sens.* **2020**, *12*, 750. [[CrossRef](#)]
34. Shafiee, S.; Lied, L.M.; Burud, I.; Dieseth, J.A.; Alsheikh, M.; Lillemo, M. Sequential forward selection and support vector regression in comparison to LASSO regression for spring wheat yield prediction based on UAV imagery. *Comput. Electron. Agric.* **2021**, *183*, 106036. [[CrossRef](#)]
35. Frank, E.; Hall, M.; Holmes, G.; Kirkby, R.; Pfahringer, B.; Witten, I.H.; Trigg, L. Weka—A Machine Learning Workbench for Data Mining. In *Data Mining and Knowledge Discovery Handbook*; Maimon, O., Rokach, L., Eds.; Springer: Boston, MA, USA, 2010; pp. 1269–1277.
36. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
37. Kohavi, R. The power of decision tables. In *Proceedings of the Machine Learning: ECML-95, Berlin/Heidelberg, Germany, 25–27 April 1995*; pp. 174–189.
38. Kaur, E.D.P.; Singh, E.P. A comparative research of rule based classification on dataset using WEKA TOOL. *Int. Res. J. Eng. Technol. (IRJET)* **2019**, *6*, 2098–2102.
39. Pham, H.T.; Awange, J.; Kuhn, M. Evaluation of Three Feature Dimension Reduction Techniques for Machine Learning-Based Crop Yield Prediction Models. *Sensors* **2022**, *22*, 6609. [[CrossRef](#)]
40. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
41. Hamza, M.; Larocque, D. An empirical comparison of ensemble methods based on classification trees. *J. Stat. Comput. Simul.* **2005**, *75*, 629–643. [[CrossRef](#)]
42. Leo, S.; De Antoni Migliorati, M.; Grace, P.R. Predicting within-field cotton yields using publicly available datasets and machine learning. *Agron. J.* **2021**, *113*, 1150–1163. [[CrossRef](#)]
43. Cutler, A.; Cutler, D.R.; Stevens, J.R. Random Forests. In *Ensemble Machine Learning: Methods and Applications*; Zhang, C., Ma, Y., Eds.; Springer: Boston, MA, USA, 2012; pp. 157–175.
44. Murtagh, F. Multilayer perceptrons for classification and regression. *Neurocomputing* **1991**, *2*, 183–197. [[CrossRef](#)]
45. Taud, H.; Mas, J.F. Multilayer Perceptron (MLP). In *Geomatic Approaches for Modeling Land Change Scenarios*; Camacho Olmedo, M.T., Paegelow, M., Mas, J.-F., Escobar, F., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 451–455.
46. 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](#)]
47. Gardner, M.W.; Dorling, S.R. Artificial neural networks (the multilayer perceptron)—A review of applications in the atmospheric sciences. *Atmos. Environ.* **1998**, *32*, 2627–2636. [[CrossRef](#)]
48. Sánchez, B.; Rasmussen, A.; Porter, J.R. Temperatures and the growth and development of maize and rice: A review. *Glob. Chang. Biol.* **2014**, *20*, 408–417. [[CrossRef](#)]
49. Ray, D.K.; Gerber, J.S.; MacDonald, G.K.; West, P.C. Climate variation explains a third of global crop yield variability. *Nat. Commun.* **2015**, *6*, 5989. [[CrossRef](#)]
50. Meng, L.; Liu, H.; L. Ustin, S.; Zhang, X. Predicting Maize Yield at the Plot Scale of Different Fertilizer Systems by Multi-Source Data and Machine Learning Methods. *Remote Sens.* **2021**, *13*, 3760. [[CrossRef](#)]
51. Reidsma, P.; Ewert, F.; Boogaard, H.; Diepen, K.v. Regional crop modelling in Europe: The impact of climatic conditions and farm characteristics on maize yields. *Agric. Syst.* **2009**, *100*, 51–60. [[CrossRef](#)]
52. Shao, R.-x.; Yu, K.-k.; Li, H.-w.; Jia, S.-j.; Yang, Q.-h.; Zhao, X.; Zhao, Y.-l.; Liu, T.-x. The effect of elevating temperature on the growth and development of reproductive organs and yield of summer maize. *J. Integr. Agric.* **2021**, *20*, 1783–1795. [[CrossRef](#)]
53. Zhang, L.; Zhang, Z.; Luo, Y.; Cao, J.; Tao, F. Combining Optical, Fluorescence, Thermal Satellite, and Environmental Data to Predict County-Level Maize Yield in China Using Machine Learning Approaches. *Remote Sens.* **2020**, *12*, 21. [[CrossRef](#)]

54. Hatfield, J.L. Increased Temperatures Have Dramatic Effects on Growth and Grain Yield of Three Maize Hybrids. *Agric. Environ. Lett.* **2016**, *1*, 150006. [[CrossRef](#)]
55. Edreira, J.I.R.; Mayer, L.I.; Otegui, M.E. Heat stress in temperate and tropical maize hybrids: Kernel growth, water relations and assimilate availability for grain filling. *Field Crops Res.* **2014**, *166*, 162–172. [[CrossRef](#)]
56. Lizaso, J.I.; Ruiz-Ramos, M.; Rodríguez, L.; Gabaldon-Leal, C.; Oliveira, J.A.; Lorite, I.J.; Sánchez, D.; García, E.; Rodríguez, A. Impact of high temperatures in maize: Phenology and yield components. *Field Crops Res.* **2018**, *216*, 129–140. [[CrossRef](#)]
57. Siatwiinda, S.M.; Supit, I.; van Hove, B.; Yerokun, O.; Ros, G.H.; de Vries, W. Climate change impacts on rainfed maize yields in Zambia under conventional and optimized crop management. *Clim. Chang.* **2021**, *167*, 39. [[CrossRef](#)]
58. Omoyo, N.N.; Wakhungu, J.; Oteng'i, S. Effects of climate variability on maize yield in the arid and semi arid lands of lower eastern Kenya. *Agric. Food Secur.* **2015**, *4*, 8. [[CrossRef](#)]
59. Adamgbe, E.M.; Ujoh, F. Effect of variability in rainfall characteristics on maize yield in Gboko, Nigeria. *J. Environ. Prot.* **2013**, *4*, 36308. [[CrossRef](#)]
60. János, N. Impact of Fertilization and Irrigation on the Correlation between the Soil Plant Analysis Development Value and Yield of Maize. *Commun. Soil Sci. Plant Anal.* **2010**, *41*, 1293–1305. [[CrossRef](#)]
61. Balogh, P.; Bujdos, Á.; Czibere, I.; Fodor, L.; Gabnai, Z.; Kovách, I.; Nagy, J.; Bai, A. Main Motivational Factors of Farmers Adopting Precision Farming in Hungary. *Agronomy* **2020**, *10*, 610. [[CrossRef](#)]
62. Cheng, M.; Wang, H.; Fan, J.; Zhang, F.; Wang, X. Effects of Soil Water Deficit at Different Growth Stages on Maize Growth, Yield, and Water Use Efficiency under Alternate Partial Root-Zone Irrigation. *Water* **2021**, *13*, 148. [[CrossRef](#)]
63. Guyader, J.; Baron, V.S.; Beauchemin, K.A. Effect of Harvesting Corn after Frost in Alberta (Canada) on Whole-Plant Yield, Nutritive Value, and Kernel Properties. *Agronomy* **2021**, *11*, 459. [[CrossRef](#)]
64. Lobell, D.B.; Hammer, G.L.; McLean, G.; Messina, C.; Roberts, M.J.; Schlenker, W. The critical role of extreme heat for maize production in the United States. *Nat. Clim. Chang.* **2013**, *3*, 497–501. [[CrossRef](#)]
65. Cedric, L.S.; Adoni, W.Y.H.; Aworka, R.; Zoueu, J.T.; Mutombo, F.K.; Krichen, M.; Kimpolo, C.L.M. Crops yield prediction based on machine learning models: Case of West African countries. *Smart Agric. Technol.* **2022**, *2*, 100049. [[CrossRef](#)]
66. Abbas, F.; Afzaal, H.; Farooque, A.A.; Tang, S. Crop Yield Prediction through Proximal Sensing and Machine Learning Algorithms. *Agronomy* **2020**, *10*, 1046. [[CrossRef](#)]
67. 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](#)]
68. Ruan, G.; Li, X.; Yuan, F.; Cammarano, D.; Ata-Ul-Karim, S.T.; Liu, X.; Tian, Y.; Zhu, Y.; Cao, W.; Cao, Q. Improving wheat yield prediction integrating proximal sensing and weather data with machine learning. *Comput. Electron. Agric.* **2022**, *195*, 106852. [[CrossRef](#)]
69. Shetty, S.A.; Padmashree, T.; Sagar, B.M.; Cauvery, N.K. Performance Analysis on Machine Learning Algorithms with Deep Learning Model for Crop Yield Prediction. In Proceedings of the Data Intelligence and Cognitive Informatics, Tirunelveli, India, 8–9 July 2021; pp. 739–750.
70. Torsoni, G.B.; de Oliveira Aparecido, L.E.; dos Santos, G.M.; Chiquitto, A.G.; da Silva Cabral Moraes, J.R.; de Souza Rolim, G. Soybean yield prediction by machine learning and climate. *Theor. Appl. Climatol.* **2023**, *151*, 1709–1725. [[CrossRef](#)]
71. Elbeltagi, A.; Srivastava, A.; Kushwaha, N.L.; Juhász, C.; Tamás, J.; Nagy, A. Meteorological Data Fusion Approach for Modeling Crop Water Productivity Based on Ensemble Machine Learning. *Water* **2023**, *15*, 30. [[CrossRef](#)]
72. Kuradusenge, M.; Hitimana, E.; Hanyurwimfura, D.; Rukundo, P.; Mtonga, K.; Mukasine, A.; Uwitonze, C.; Ngabonziza, J.; Uwamahoro, A. Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agriculture* **2023**, *13*, 225. [[CrossRef](#)]
73. Keerthana, M.; Meghana, K.J.M.; Pravallika, S.; Kavitha, M. An Ensemble Algorithm for Crop Yield Prediction. In Proceedings of the 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), Tirunelveli, India, 4–6 February 2021; pp. 963–970.
74. Jeong, J.H.; Resop, J.P.; Mueller, N.D.; Fleisher, D.H.; Yun, K.; Butler, E.E.; Timlin, D.J.; Shim, K.-M.; Gerber, J.S.; Reddy, V.R.; et al. Random Forests for Global and Regional Crop Yield Predictions. *PLoS ONE* **2016**, *11*, e0156571. [[CrossRef](#)]
75. Cheng, M.; Penuelas, J.; McCabe, M.F.; Atzberger, C.; Jiao, X.; Wu, W.; Jin, X. Combining multi-indicators with machine-learning algorithms for maize yield early prediction at the county-level in China. *Agric. For. Meteorol.* **2022**, *323*, 109057. [[CrossRef](#)]
76. 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](#)]
77. Ahmed, S. A Software Framework for Predicting the Maize Yield Using Modified Multi-Layer Perceptron. *Sustainability* **2023**, *15*, 3017. [[CrossRef](#)]
78. Paswan, R.P.; Begum, S.A. ANN for prediction of Area and Production of Maize crop for Upper Brahmaputra Valley Zone of Assam. In Proceedings of the 2014 IEEE International Advance Computing Conference (IACC), New Delhi, India, 21–22 February 2014; pp. 1286–1295.
79. Kaul, M.; Hill, R.L.; Walthall, C. Artificial neural networks for corn and soybean yield prediction. *Agric. Syst.* **2005**, *85*, 1–18. [[CrossRef](#)]

80. 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](#)]
81. Bhojani, S.H.; Bhatt, N. Wheat crop yield prediction using new activation functions in neural network. *Neural Comput. Appl.* **2020**, *32*, 13941–13951. [[CrossRef](#)]
82. Hara, P.; Piekutowska, M.; Niedbała, G. Selection of Independent Variables for Crop Yield Prediction Using Artificial Neural Network Models with Remote Sensing Data. *Land* **2021**, *10*, 609. [[CrossRef](#)]
83. Abraham, E.R.; Mendes dos Reis, J.G.; Vendrametto, O.; Oliveira Costa Neto, P.L.d.; Carlo Toloi, R.; Souza, A.E.d.; Oliveira Morais, M.d. Time Series Prediction with Artificial Neural Networks: An Analysis Using Brazilian Soybean Production. *Agriculture* **2020**, *10*, 475. [[CrossRef](#)]
84. Son, N.-T.; Chen, C.-F.; Chen, C.-R.; Guo, H.-Y.; Cheng, Y.-S.; Chen, S.-L.; Lin, H.-S.; Chen, S.-H. Machine learning approaches for rice crop yield predictions using time-series satellite data in Taiwan. *Int. J. Remote Sens.* **2020**, *41*, 7868–7888. [[CrossRef](#)]
85. Li, L.; Wang, B.; Feng, P.; Li Liu, D.; He, Q.; Zhang, Y.; Wang, Y.; Li, S.; Lu, X.; Yue, C. Developing machine learning models with multi-source environmental data to predict wheat yield in China. *Comput. Electron. Agric.* **2022**, *194*, 106790. [[CrossRef](#)]
86. Chen, X.; Feng, L.; Yao, R.; Wu, X.; Sun, J.; Gong, W. Prediction of Maize Yield at the City Level in China Using Multi-Source Data. *Remote Sens.* **2021**, *13*, 146. [[CrossRef](#)]
87. Ngie, A.; Ahmed, F. Estimation of Maize grain yield using multispectral satellite data sets (SPOT 5) and the random forest algorithm. *S. Afr. J. Geomat.* **2018**, *7*, 11–30. [[CrossRef](#)]
88. Kayad, A.; Sozzi, M.; Gatto, S.; Marinello, F.; Pirotti, F. Monitoring Within-Field Variability of Corn Yield using Sentinel-2 and Machine Learning Techniques. *Remote Sens.* **2019**, *11*, 2873. [[CrossRef](#)]
89. Zhu, X.; Guo, R.; Liu, T.; Xu, K. Crop Yield Prediction Based on Agrometeorological Indexes and Remote Sensing Data. *Remote Sens.* **2021**, *13*, 2016. [[CrossRef](#)]
90. Nagy, A.; Fehér, J.; Tamás, J. Wheat and maize yield forecasting for the Tisza river catchment using MODIS NDVI time series and reported crop statistics. *Comput. Electron. Agric.* **2018**, *151*, 41–49. [[CrossRef](#)]

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