

Article

Towards the Modeling and Prediction of the Yield of Oilseed Crops: A Multi-Machine Learning Approach

Mahdieh Parsaeian ¹, Mohammad Rahimi ², Abbas Rohani ^{2,*}  and Shaneka S. Lawson ³

¹ Department of Agronomy and Plant Breeding, Shahrood University of Technology, Shahrood 3619995161, Iran

² Department of Biosystems Engineering, Ferdowsi University of Mashhad, Mashhad 9177948974, Iran

³ USDA Forest Service, Northern Research Station, Hardwood Tree Improvement and Regeneration Center (HTIRC), Department of Forestry and Natural Resources, Purdue University, 715 West State Street, West Lafayette, IN 47906, USA

* Correspondence: arohani@um.ac.ir

Abstract: Crop seed yield modeling and prediction can act as a key approach in the precision agriculture industry, enabling the reliable assessment of the effectiveness of agro-traits. Here, multiple machine learning (ML) techniques are employed to predict sesame (*Sesamum indicum* L.) seed yields (SSY) using agro-morphological features. Various ML models were applied, coupled with the PCA (principal component analysis) method to compare them with the original ML models, in order to evaluate the prediction efficiency. The Gaussian process regression (GPR) and radial basis function neural network (RBF-NN) models exhibited the most accurate SSY predictions, with determination coefficients, or R^2 values, of 0.99 and 0.91, respectfully. The root-mean-square error (RMSE) obtained using the ML models ranged between 0 and 0.30 t/ha (metric tons/hectare) for the varied modeling process phases. The estimation of the sesame seed yield with the coupled PCA-ML models improved the performance accuracy. According to the k-fold process, we utilized the datasets with the lowest error rates to ensure the continued accuracy of the GPR and RBF models. The sensitivity analysis revealed that the capsule number per plant (CPP), seed number per capsule (SPC), and 1000-seed weight (TSW) were the most significant seed yield determinants.

Keywords: agro-morphological; data-driven; machine learning; seed yield; sensitivity analysis



Citation: Parsaeian, M.; Rahimi, M.; Rohani, A.; Lawson, S.S. Towards the Modeling and Prediction of the Yield of Oilseed Crops: A Multi-Machine Learning Approach. *Agriculture* **2022**, *12*, 1739. <https://doi.org/10.3390/agriculture12101739>

Academic Editor: Yanqun Zhang

Received: 27 August 2022

Accepted: 18 October 2022

Published: 21 October 2022

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



Copyright: © 2022 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

The continuous spread of industry, along with rapid resource depletion, has increased the demand for green energy sources. Fossil fuels (e.g., oil, coal, and natural gas) are non-renewable resources and, despite being the leading contributors to rising CO₂ emission levels, have long been utilized by industries as fuel [1–3]. Oilseeds are regarded as one of the most important energy sources, with diverse industrial and medicinal applications. Therefore, the precise prediction of the crop yield is a principal objective for agricultural and industrial applications [4,5]. Forecasting the crop yields prior to harvest can help one to identify optimal reaping times and alleviate the concerns of farmers regarding field conditions and management [6,7]. As a result, it is critical to enhance the planting methods for oilseed species and develop new cultivars with higher potential yields. Sesame (*Sesamum indicum* L.) is one of the oldest oilseeds, with its nutritious seeds containing oil (34.4–63.2%), proteins (17–32%), minerals, and fat-soluble vitamins [8,9]. Sesame oil is the most stable and high-quality edible oil due to its unique combination of fatty acids and natural antioxidants. However, little research exists regarding the planting and development of adaptable, high-yield cultivars [10–12]. As a crucial breeding objective, the yield is a complex, quantitative, polygenic trait that is primarily influenced by several factors underpinning production. The phenotypic representation of this trait is typically impacted by the environment and environment–genotype interaction. Thus, it is seldom heritable, and the efficacy and efficiency of long-term direct selection for this trait are restricted. [13].

In contrast, the selection of seed-yield-related traits that are heritable is a promising method that may be used to improve the seed crop yield. These traits are relatively insensitive to the environment and often highly heritable [14–16]. Yield components can indirectly affect the seed yield through their positive or negative interactions. Thus, deducing the relationships between the seed yield and agro-traits is regarded as an effective approach to trait enhancement.

Statistical modeling applications have been widely used to explain the relationships between the morphological and agronomic traits affecting the sesame seed yield. Other yield prediction methods, such as quadratic, pure quadratic, interaction (2FI), and polynomial methods, have previously been used for cotton, maize, and wheat crops. The best regression model for this study was chosen based on the values of the assessment criteria [17]. In another study, regression analyses were adapted for the survey of major environmental factors and their impacts on the crop yield. Yield predictions were believed to provide substantial benefits to farmers while reducing crop loss and increasing earnings [18]. Alternatively, the multiple linear regression (MLR) technique was employed in the East Godavari district of Andhra Pradesh in India to predict crop yields. Those findings, in comparison with the dataset currently available, can aid in efforts to evaluate the efficacy of the proposed technique [19]. A regression model was also manipulated to reveal the relative importance of agronomic traits and the genetic correlations with the sesame seed yield. The model conveyed data to support the notion that the CPP has stronger direct and positive effects on the seed yield than most other traits [20]. Another study heralded the significance of the CPP and deemed it important for yield selections [21], while other studies reported that greater plant heights, combined with a higher CPP, could increase the overall sesame seed yield [22,23]. Similarly, the seeds per capsule (SPC), thousand-seed weight (TSW), and the number of capsules per branch (CPB) were positively and significantly correlated with the seed yield [22,24]. Additional studies in this area applied multiple linear regression (MLR) models developed to assess crop yield traits. Independent variables (inputs) affecting the seed yield were identified and considered; however, the CPP was the foremost variable required for the best results [25]. Similar results were achieved when fitting a predictor equation for the seed yield [26,27]. Although traditional statistical methods (i.e., regression analyses) are widely used to derive plant seed yield prediction equations, assumptions, such as the normality of the dependent variables, homogeneity of the error variance, and inefficient representation of the nature of complex and nonlinear relations in empirical phenomena, represent substantial drawbacks [28].

Machine learning (ML) techniques have attracted extensive attention because they can easily be used in fields such as agriculture and chemical and energy sciences for a variety of applications [29–35]. Consequently, agronomists have shifted towards machine learning methods such as artificial neural networks (ANNs) and Gaussian process regression (GPR) models in recent years [36–40]. ML models are especially effective in agricultural fields and have been used for product image processing [41], the separation of weeds and vegetative cover in remote sensing [42], the prediction of solar radiation [43], flood forecasting [44], hydrogen storage on bio-carbon [45], biomass estimation [46], CO₂ capture [47], and the estimation of soil erosion rates [48]. As shown in Table 1, numerous studies have expounded upon the usefulness of ML in investigations of seed and crop yields. Moreover, predictions of agro-product constituents, such as oil or nitrogen contents or disease diagnosis and plant classifications, are most often accomplished with ML models. These intelligent models use numerous interconnected processing elements to solve problems and can be modified to perform specific functions, including pattern identification, data classification, and the prediction and modeling of processes through a reliable learning process [49,50]. ANNs are characterized by their suitable error tolerance, direct learning from data, and lack of a need for statistical quantity estimations [51,52]. The predictions of the output correspond to a set of inputs, where parameter relationships serve different functions based on the study goals [16]. In the agriculture field, ML is most often tasked with investigating multi-objective concerns, such as crop yield estimation and quality control. A selection of

agricultural research studies devoted to crop yields, plant classification, seed assessments, and crop quality control, in which ML was incorporated, is illustrated in Figure 1. Radial basis function neural networks (RBF-NN) and regression models have been adapted for the prediction of the tree trunk volume. The RBF neural network has been operationally more reliable than regression models in completing this task [53].

Table 1. A selection of previous studies using ML to advance agricultural crop research.

Application	Performance Prediction	Model	RMSE (R ²) *	Ref.
Seed and Crop Yield	Prediction of oilseed rape yield with alternative planting styles and varied nitrogen fertilizer applications.	SVR ANN PLSR	-	[54]
	Estimation of soybean seed yield using collected multispectral images for predictions.	MLP	-	[55]
	Incorporation of multi-qualitative and quantitative features for the estimation of wheat yields.	ANN	-	[56]
	ML model comprised of high-dimensional phenotypic trait data to carry out in-season seed yield predictions.	RF	- (0.83)	[57]
	Ten agro-morphological and phenological traits (plant height, number of branches per plant, number of capsules per plant, number of days to flowering, number of days to maturity, thousand-seed weight, etc.) were used as the basis for a predictive seed yield model.	ANN MLP RBF PCA	0.87 (0.92)	[16,58]
	Prediction of crop yields in mustard and potato with models using soil elemental properties, physicochemical features, pH, electrical conductivity, organic carbon, and others for training and test datasets.	ANN SVR KNN	- 4.62 (0.72)	[59,60]
	Prediction of corn crop yield by careful climate change factor (temperature and moisture) evaluation to compile an impact assessment of corn fields.	ANN	1.5	[61]
	ML to provide predictive estimates of the maize crop yield using topography, land use, soil data, and multiple other parameters.	-	(0.96)	[62]
	Yield predictions of rice paddies using climate-based factors (rainfall, morning and evening relative humidity, minimum and maximum temperature).	ANN	31	[63]
	Utilization of fertilizer volume in tandem with general atmospheric conditions to predict maize yields.	ANN MR	30	[64]
	Predictions of rice paddy yields based on environmental features (area, number of open wells, tanks, maximum temperature, etc.) as independent variables.	ANN MLR SVR RF	0.05–0.1 (0.8)	[65]
	Constructing several distinct ML models to predict winter rapeseed yield at specific timepoints from six agro-morphological traits (oil and protein content, seed yield, oil and protein yield, and thousand-seed weight) as inputs.	ANN RF	- (0.944)	[66,67]
	Examination of micro-topographic attributes related to growth in agronomic crops based on analyses of vegetation indices, lidar derivatives, and crop type.	ANN	-	[68]
	Investigation of available water holding capacity of soil coupled with climate data, used to estimate the average wheat yield within a region.	ANN	-	[69]
	Cotton lint yield derived from a remote sensing ANN model evaluating eight phenological crop indices.	ANN	-	[70]

Table 1. Cont.

Application	Performance Prediction	Model	RMSE (R ²) *	Ref.
	Six ML algorithms applied to predict the cotton yield by climate and management parameters.	ANN, RF, etc.	(0.51)	[49]
	Predicted and optimized the corn yield by ML technique.	-	9	[71]
	Some seed and crop yields, such as maize, sorghum, and groundnut, modeled by climate data.	-	-	[72]
	NN-based ML applied to predict the chemical specification (fatty acid) of rubber seed oil	ANN and ANFIS	(0.69 to 0.99)	[73]
	Prediction of seed yield with ML and coupled PCA-ML models accompanied by sensitivity analysis based on agro-morphological features of sesame plants.	MLR PCA RBF GPR	0.00–0.36 (0.88–0.99)	<i>This work</i>
Seed classification	14 categories of seeds and date fruits are classified by utilizing deep learning.	CNNs	(0.99)	[74,75]
	ML applied to classify various seeds by using simple architecture and memory characteristics.	CNNs	(0.98)	[76]
Nitrogen and Oil Concentration	Nitrogen prediction of oilseed rape leaves based on ten spectral features from both barley and oilseed rape.	ANN	0.30 (0.9)	[77,78]
	The merging of bio-physiochemical and spectral features in leaves for further in-depth studies.	GP	2.2–5.8	[79]
	Prediction of sesame oil content from eighteen agro-morphological and phenological traits using ML in efforts to prevent marginal effects.	ANN MLR PCA	0.56 (0.86)	[80]
Disease and Quality Diagnoses for use in Classification	Integration of ML models coupled with hyperspectral imaging to detect disease in pre-symptomatic tobacco plants.	LS-SVM	-	[81]
	Conducting an oilseed disease analysis with directed surveys of ten common oilseed disease classes.	DT RF MLP	-	[82]
	Investigation of physicochemical properties (fatty acid and mineral profiles) and additional physical attributes of six sunflower varieties for use in classification, grading, and quality assessment studies.	SVM, RF, MLR, etc.	0.21 (0.81)	[83]

* Column indicates the best RMSE value, or R² value (in parenthesis), obtained in the referenced study. Artificial neural network (ANN); adaptive neuro-fuzzy inference system (ANFIS); convolutional neural networks (CNNs); DT (decision tree); GP (Gaussian process); k-NN (k-nearest neighbor); LS-SVM (least-squares support vector machines); PLSR (partial least-squares regression); RF (random forest); SVR (support vector regression); PCA (principal component analysis); PHN (pruning hidden nodes).

Similarly, the RBF-NN is more efficient than the multilayer perceptron neural (MLP) network in the prediction of rice yields in terms of its training time, precision, and the number of neurons in the hidden layer [84]. Efficiency comparisons of the RBF and MLP neural networks with the MLR model indicated that ANN models more accurately estimated the biological and grain yield in barley [85]. Earlier works corroborated the greater ability of ANNs to predict the wheat performance and to map and determine rice yields [69,86]. Moreover, Gad and El-Ahmady applied PCA to predict the black seed oil yield, achieving a correlation coefficient of 0.997 [87]. GPR is also used for yield prediction in the case of agricultural products. Applying a set of random variables, GPR is capable of solving nonlinear problems using a novel data mining method [88]. Previous studies on MLR models often focused on model evaluation without checking their generalizability. To overcome such drawbacks, this study aimed to evaluate the ability of certain well-known machine learning models (e.g., ANNs and GPR) and coupled models, such as

PCA-MLs, to estimate plant yields, a gap currently found in the literature. As such, the main objectives of the present work are to estimate the agronomical yield of sesame using ML and coupled PCA-ML models to predict the SSY and to compare the resultant data from each model. The primary motivation for, and contributions of, this study include: (i) the use of the principal component analysis (PCA) approach to simplify the calculations and reduce the number of sesame production input variables, (ii) to assess the ML and PAC-ML models' generalizability and identify an optimal training and test dataset utilizing the k-fold approach, (iii) to employ MLR training and testing procedures to provide a comparison with the ML models, (iv) to predict sesame yields with the RBF-NN and GPR models for their comparison with the MLR model outputs, and (v) to apply the sensitivity analysis approach in order to uncover the features that are essential for the sesame seed yield. This research, predicting the sesame oil yield, offers the following contributions:

- Four ML models were employed to aid in predicting the sesame seed yield (SSY).
- Coupled PCA-ML models were used for in-depth predictions of SSY for the first time.
- The use of the GPR, RBF, and MLR models led to a greater accuracy of the SSY predictions.
- The primary agro-morphological features for predicting the SSY were revealed through a sensitivity analysis.

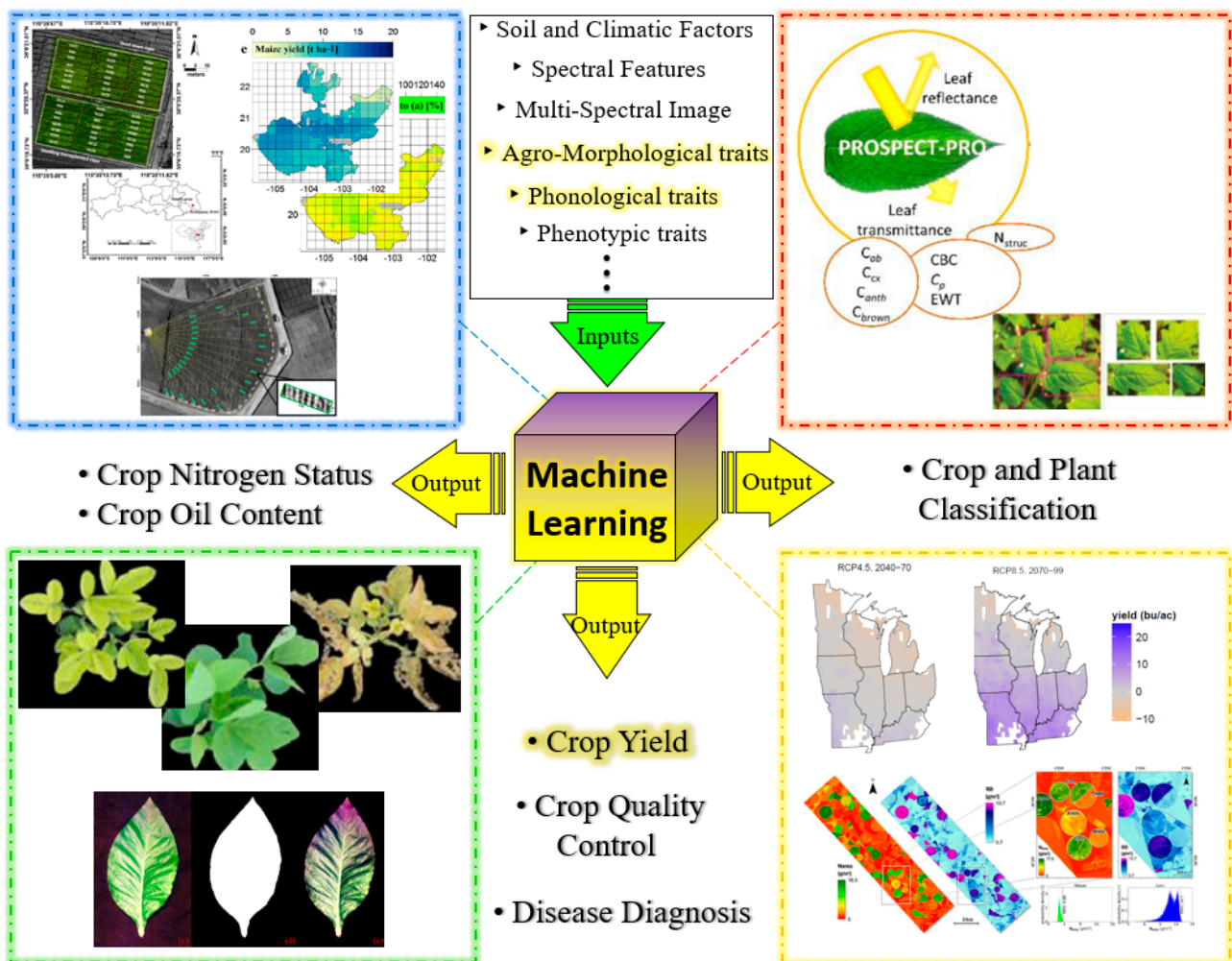


Figure 1. Overview of the potential agricultural research questions that have been investigated using ML modeling methods [54,55,62,79,81,89–92].

2. Material and Method

2.1. Field Experiments

In this study, 135 sesame genotypes were derived from five genotypes (selected from nine diverse genotypes) representing various sesame growing zones within Iran (Varamin 2822, Borazjan 1, Darab 1, Tn₂₄₀, and Tn₂₃₄). The selected genotypes were produced from various Iranian landraces planted on a research farm at the Shahrud University of Technology (36.39° N, 54.94° E) during 2018–2019. Each experimental plot consisted of two 1.5 m rows with 50 cm spacing between rows and 7 cm spacing between plants. The plots were fertilized with 80 kg ha⁻¹ N and 100 kg ha⁻¹ P before sowing and 40 kg ha⁻¹ N upon flower initiation. The crops were grown in a clay loam Typic Haplargid aridisol at pH 7.5 with 1% soil organic matter. Agro-morphological data were collected from the F1 and F2 progenies (81 F1 and 45 F2), arranged in a randomized complete block design (RCBD). As illustrated in Figure 2, the provided datasets were applied to construct the ML models that were used to predict the seed yields and relative importance of input features. A total of 135 sesame genotypes were agronomically investigated in the form of an RCBD, with 3 replications. A total of 10 random plants were measured from each experimental unit and the average was considered as the result of that repetition. Therefore, a total of 135 * 3 = 405 data were obtained. Based on the Grubbs test, 27 data were identified as outliers and were removed from the dataset. Therefore, 378 data remained for the final dataset. Next, 80% and 20% of the dataset, including 302 and 76 data, were used for the training and testing of machine learning methods, respectively.

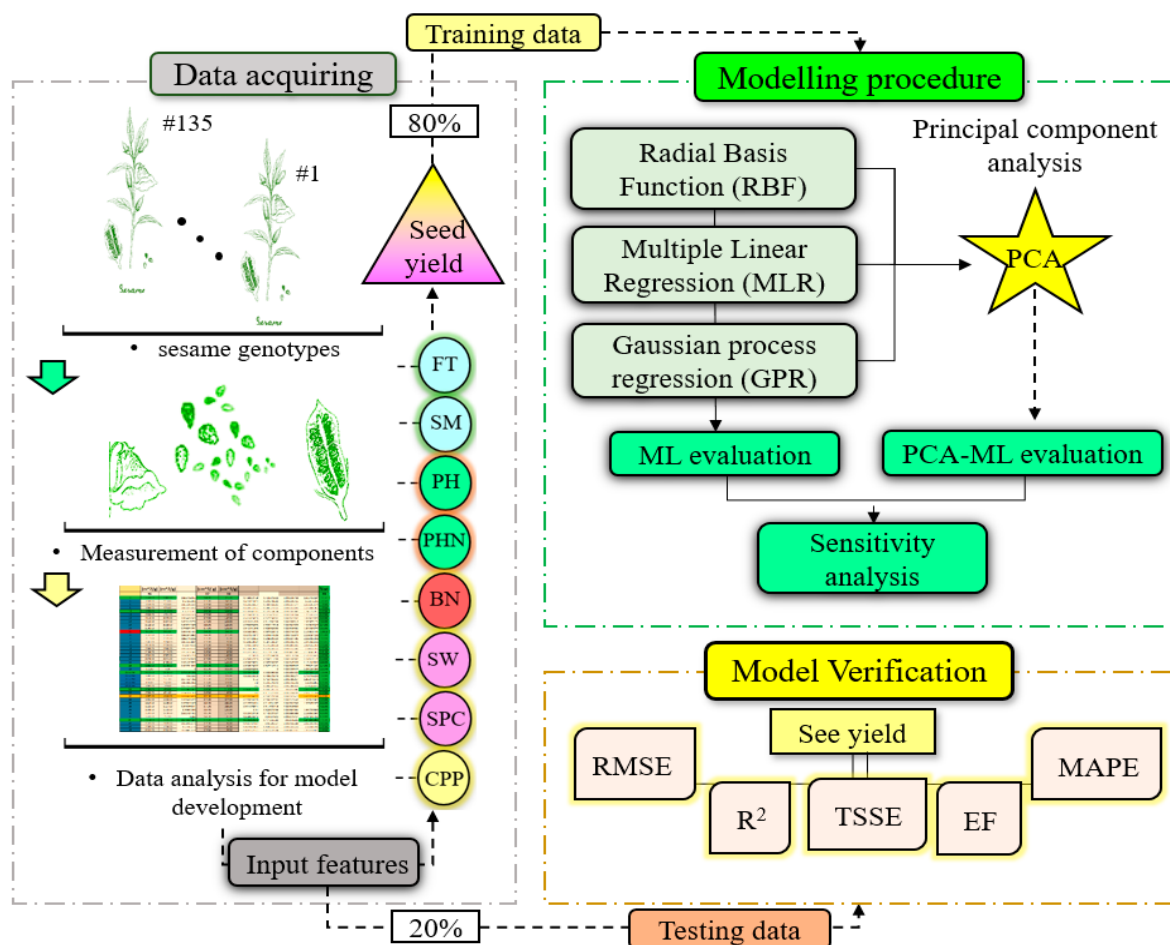


Figure 2. The schematic diagram of the dataset provision, modeling, and model verification processes for the seed yield prediction.

The experiment was repeated in triplicate, and the agro-morphological traits (independent variables) are illustrated in Table 2. These variables include the flowering time at 10% (FT-10) and 100% (FT-100), seed maturity (SM), plant height (PH), the height of the first fruit-bearing node (PHN), number of fruit-bearing branches (BN), capsule number per plant (CPP), seed number per capsule (SPC), and 1000-seed weight (TSW). FT 10% means the time by which approximately 10% of plant had flowered, while FT 100% is the time by which the plant had completely flowered and grown. These input features were measured on 10 plants randomly selected from each sampled plot. The remaining plants were used after eliminating marginal effects to determine the output as the sesame seed yield (SSY). Input variables are defined as x_1 to x_9 , while (y), the sesame seed yield, is designated as the output. A summary of the measurements obtained is presented in Table 2. Visual data were collected daily from each plot for FT10, FT100, and SM. The SSY data were obtained by harvesting two rows from the middle of the experimental plot.

Table 2. Summary of the measured statistical parameters for the agronomic trait data.

Variables	Symbols	Input	Min	Max	Mean	Stdv
Flowering time 10% (days)	FT10	x1	41	55	47.25	2.57
Flowering time 100% (days)	FT100	x2	41	60	52.17	2.73
Seed maturity (days)	SM	x3	102	155	130.63	13.05
Plant height (cm)	PH	x4	100.86	199.2	143.45	16.14
Plant height to first fruiting node (cm)	PHN	x5	26.15	87.92	58.26	10.65
Capsule number per plant	CPP	x6	24.68	99.21	50.83	11.94
Thousand-seed weight (g)	TSW	x7	1.98	4.22	3.41	0.36
Seed number per capsule (g)	SPC	x8	27.74	102.61	54.52	13.47
Branch number	BN	x9	0	5.3	1.84	1.08
Seed yield of sesame (t/ha)	SSY	y	1.19	4.18	2.52	0.58

2.2. Machine Learning Algorithms and Models

2.2.1. Multiple Linear Regression (MLR) Models

The study assessed four MLR models, including linear, interaction (2FI), reduced quadratic, and quadratic models, as can be seen in Equations (1)–(4), respectively. [93]. The ANOVA regression coefficients were determined using MATLAB software 2019a (MathWorks Inc., Natick, MA, USA):

$$y = \beta_0 + \sum_{i=1}^9 \beta_i x_i + \varepsilon \quad (1)$$

$$y = \beta_0 + \sum_{i=1}^9 \beta_i x_i + \sum_{i=1}^9 \sum_{j=i+1}^9 \beta_{ij} x_{ij} + \varepsilon \quad (2)$$

$$y = \beta_0 + \sum_{i=1}^9 \beta_i x_i + \sum_{i=1}^9 \beta_{ii} x_{ii} + \varepsilon \quad (3)$$

$$y = \beta_0 + \sum_{i=1}^9 \beta_i x_i + \sum_{i=1}^9 \sum_{j=i+1}^9 \beta_{ij} x_{ij} + \sum_{i=1}^9 \beta_{ii} x_{ii} + \varepsilon \quad (4)$$

where y is the crop yield (t/ha^{-1}), x represents the independent variables (inputs), β_0 is the intercept, β_i is the linear regression coefficient, β_{ij} is the interaction regression coefficient, and β_{ii} is the quadratic regression coefficient.

2.2.2. Principal Component Analysis (PCA)

PCA is a tool for reducing model calculations and input vector dimensions. The basis of PCA is the transformation of vast amounts of model input data into a smaller set of new variables (principal components) with lower autocorrelations [94,95]. The procedure is as follows:

$$X = X - \bar{X} \quad (5)$$

$$C = \frac{1}{n}XX^T \quad (6)$$

$$D = V^{-1}CV \quad (7)$$

$$Z = XV \quad (8)$$

where X is the input vector matrix, C is the covariance matrix, V is the eigenvector of C , D is the vector of the eigenvalues of C , and Z is the eigenvector of X .

2.2.3. Gaussian Process Regression (GPR) Model

The GPR model is a non-parametric, kernel-based probabilistic regression framework, which infers functions from a set of training data $D = \{(y_n, x_n), n = 1, 2, 3, \dots, N\}$ of N vector input pairs, $x_n \in \mathbb{R}^L$, and output y_n within a noisy scalar field. Effective for smaller datasets, this Bayesian model effectively generalizes the output distribution in unrevealed input zones. The output noise, or model uncertainty, is often caused by external factors unrelated to x , such as observation errors. The model noise assumption is zero-mean and is defined as:

$$y = f(x) + \varepsilon, \quad \varepsilon \gg N(0, \sigma_{noise}^2) \quad (9)$$

where σ_{noise}^2 equals the noise variance. GPR utilizes the Gaussian process (GP) to describe a latent variable function, referred to as f , with x used to describe index-related latent variables in a finite collection $\{f(x_1), \dots, f(x_k)\}$ where the aforementioned indices constitute a consistent normal or “Gaussian” distribution. This allows for nonlinear regression between latent variable pairs. There are several advantages to GPR, such as the ability to estimate the model uncertainty and the ability to use the estimations to specify function types. The mean function $m(x)$ and $k(x, x')$, which equals the kernel, or covariance, function, where E equals expectation, are defined as:

$$m(x) = E[f(x)], \quad k(x, x') = E[(f(x) - m(x))(f(x') - m(x')))] \quad (10)$$

Typically defined as either zero or the dataset mean, the mean function is typically constant. The mean function is significant only with respect to the average behavior of the model over time, while the covariance function, a more comprehensive value, includes all the procedure observations. The covariance function most often uses a hierarchical model, where covariance parameters, called hyperparameters, define the distribution $f(x)$. The squared exponential covariance function employed to generate a smooth path is defined as:

$$K(x, x') = q_1 \exp(-(\|x - x'\|)^2 / (2q_2)) \quad (11)$$

A stationary covariance function and Euclidian norm, $\|\cdot\|$, is a function of $x - x'$ and is invariant to changes in the input or x -space. With an increase in the x -space distance, or the space between x and x' , the decay in covariance escalates exceptionally quickly. This implies that correlations between $f(x)$ and $f(x')$ are negligible. The hyperparameter q_1 specifies the maximum permissible covariance, while q_2 defines the rate of decay as the correlation distance increases. The covariance matrix represents the relatedness of one observation to another based on a set of kernel parameters. It can be defined as:

$$K(x_i, x_j) = \sigma_f^2 \exp(-(x_i - x_j)^2 / 2l^2) + \sigma_n^2 \delta(x_i, x_j) \quad (12)$$

where σ_f^2 is the maximum acceptable covariance, l is the covariance matrix length parameter, and $\delta(x_i, x_j)$ is the Kronecker delta function. The covariance matrix is assessed during the GPR training process, and then the training dataset output is estimated.

2.2.4. Radial Basis Function (RBF) Neural Network

The RBF model is a flexible feed-forward network that is able to automatically predict and classify new output patterns after the training phase [96,97]. The structure of the RBF

model applied in the present study is illustrated in Figure 3. The independent variables serve as first layer inputs while second (hidden) layer inputs are subjected to the nonlinear activator function $\phi(r)$ before all of them are tallied together in the third or output layer. The RBF is effective in approximating non-linear input–output mapping and has a strong tolerance to input noise. The optimum values of the matrix for the weight (W) and other model parameters are acquired during the training stage by minimizing the sum of the squared errors (SSE). The RBF model output, (y), is defined as:

$$y(x) = \sum_{k=1}^m w_k \phi(\|x - c_k\|) \quad (13)$$

where w_k is the weight of the linking of the k th neuron from the hidden layer to the output layer and c_k is the pre-pattern center of the k th neuron from the hidden layer. The Gaussian radial basis function is denoted as:

$$\phi(r) = e^{-\frac{r^2}{2\sigma^2}} \quad (14)$$

where r is the distance between the input and pattern center (c) and σ is a parameter controlling the smoothness of the interpolation function [98,99].

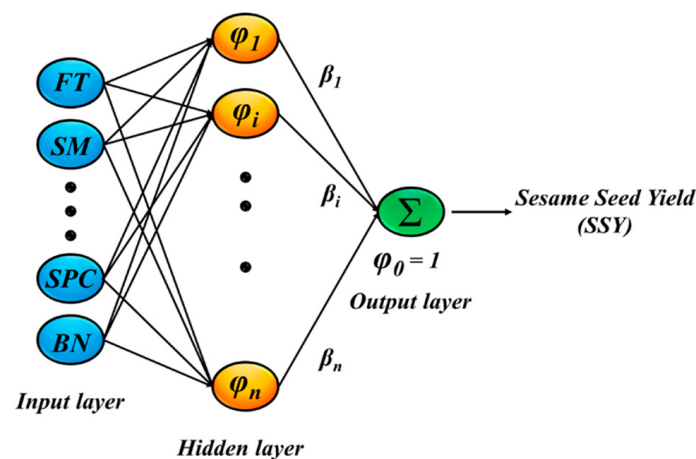


Figure 3. The RBF neural network structure applied in the present study.

2.3. K-Fold Cross-Validation

K-fold cross-validation is used as a reliable approach to ruling out model bias. To train MLR, GPR, and RBF models, the dataset is randomly divided into subsets for the training and testing phases using 80% and 20% of the total dataset. Since the input data partitioning is performed randomly, the model gives different results for each training and testing run. The types of cross-validation differ, including k-fold cross-validation, $K \times 2$ cross-validation, leave-one-out cross-validation, repeated random subsampling validation, etc. [100]. Cross-validation involves repeated random subsampling procedures, where no overlapping occurs between the test datasets. During the process, the learning set is divided into equally sized k subgroups. The “fold” refers to the number of resulting subsamples. Subsequently, one of the subsamples is designated as k and used as the test, or validation, dataset, while the remaining $k - 1$ subsamples are utilized as the training data. The first subsample selected becomes the first fold and serves as a validation sample $D_{val, 1}$ while successive subsamples serve as the training set $D_{train, 1}$. The ensuing result with the least error is named E_i . The outcomes for each k -fold are averaged or combined to obtain a single estimation. The turning parameter for k -fold cross-validation is defined [101]. Each subgroup, or $i = 1, 2, 3, \dots, k$, helps to establish the fit model with γ or $k - 1$ parts. Finally,

$\alpha^{-k}(\gamma)$ is combined with the additional computation k th to identify the prediction error and is shown as:

$$E_k(\gamma) = \sum_{i \in kth \text{ part}} [y_i - x_i \alpha^{-k}(\gamma)]^2 \quad (15)$$

This procedure continues for numerous γ cycles, and the value of γ , displaying the smallest error, is selected.

2.4. Model Assessment Criteria

Models are assessed in the training and testing steps by multiple criteria, including the mean absolute percentage error (*MAPE*), root-mean-square error (*RMSE*), efficiency factor (*EF*), and total sum squared error (*TSSE*). The linear relationship between the actual (y) and predicted values ($\hat{Y}$), including their coefficients of determination (R^2), are defined as [102]:

$$MAPE = \left(\sum_{r=1}^N \frac{|Y_r - \hat{Y}_p|}{Y_r} \right) / N \quad (16)$$

$$RMSE = \sqrt{\left(\sum_{r=1}^N |Y_r - \hat{Y}_p|^2 \right) / (N - 1)} \quad (17)$$

$$TSSE = \sum_{r=1}^N |Y_r - \hat{Y}_p|^2 \quad (18)$$

$$EF = 1 - \left(\sum_{j=1}^n (y_j - \hat{y}_j)^2 / \sum_{j=1}^n (y_j - \bar{y}_j)^2 \right) \quad (19)$$

$$R^2 = 1 - \left(\sum_1^N (Y_r - Y_p)^2 / \sum_1^N (Y_r - \hat{Y}_r)^2 \right) \quad (20)$$

where Y_r and Y_p represent the real and predicted values, and $\hat{Y}_r$ indicates the average of the real values. The *MAPE*, *RMSE*, and *TSSE* values closest to zero indicate the best ML model performance and provide accurate predictions with acceptable estimation errors. The *EF* values increase as the model approaches an optimal state. For an additional validation, the line spanning the actual and predicted values (see Section 3.3) designates the best result, which is achieved when the slope and intercept approach 1 and 0, respectively, and the coefficient of determination (R^2) nears 1.

3. Results and Discussions

3.1. Primary Statistical Analysis of Datasets

The generated datasets created inputs for the analysis of variance (ANOVA) using a general linear model. The ML models were configured using the QNet v2000 software package (QNet Ltd., Hong Kong, China) to define the SSY as the output and the other agromorphological traits as inputs. The ML models were trained and tested using 135 samples from the study area. The corresponding statistical indices for each variable are illustrated in Table 2.

The results of the Pearson correlation analysis applied to the nine independent variables are illustrated in Table 3. Almost all variables obtained a significant level in the range of ~0.01.

Table 3. The measured correlation and significant level of independent variables.

Sample 1	Sample 2	Correlation	p-Value	Sample 1	Sample 2	Correlation	p-Value
FT100	FT10	0.87	0.00	TSW	PH	0.13	0.02
SM	FT10	0.53	0.00	TSW	PHN	0.15	0.01
SM	FT100	0.54	0.00	TSW	CPP	0.13	0.01
PH	FT10	0.24	0.00	SPC	FT10	−0.17	0.00
PH	FT100	0.29	0.00	SPC	FT100	−0.10	0.07
PH	SM	0.20	0.00	SPC	SM	−0.17	0.00
PHN	FT10	0.38	0.00	SPC	PH	0.22	0.00
PHN	FT100	0.40	0.00	SPC	PHN	0.27	0.05
PHN	SM	0.34	0.00	SPC	CPP	−0.41	0.00
PHN	PH	0.65	0.00	SPC	TSW	−0.41	0.00
CPP	FT10	0.18	0.00	BN	FT10	0.17	0.13
CPP	FT100	0.13	0.01	BN	FT100	0.75	0.00
CPP	SM	0.09	0.09	BN	SM	0.55	0.03
CPP	PH	0.13	0.01	BN	PH	0.41	0.00
CPP	PHN	−0.03	0.53	BN	PHN	0.26	0.09
TSW	FT10	0.31	0.00	BN	CPP	0.48	0.01
TSW	FT100	0.28	0.00	BN	TSW	0.31	0.15
TSW	SM	0.54	0.00	BN	SPC	0.22	0.00

3.2. Statistical Processes of PCA and MLR Algorithms

The present study utilized nine agro-morphological traits to estimate the yield of sesame seeds. PCA was employed to improve the speed, reduce the calculation complexities, and decrease the number of independent variables by presenting the principal components. The obtained five principal components (PCs), as shown in Figure 4a, explained 98.99% of the variability (the total variance). Therefore, only five traits, including the pH (x4), SPC (x8), CPP (x6), SM (x3), and PHN (x5), were needed as principal components. Figure 4b shows that two variables explained 69.50% of the total variance and were more influential than the other variables as principal components. The morphological traits of 250 sesame genotypes have been evaluated by PCA. Shim et al. (2016) reported the five components accounted for 29%, 16%, 14%, 13%, and 10% of the morphological trait variance. The primary traits of the first and second components were FT, SM, and SPC [103]. Furthermore, Baraki et al. (2015) used a group of 30 African sesame genotypes to construct a PCA. Three principal components explained 88.49% of the variance in the measured agronomic traits. Their study concluded that the seed yield and oil percentage have the greatest influence on the principal component formation [104]. Ismaila and Usman (2014) also reported on three components analyzed by PCA, indicating that those three components accounted for 86.73% of the variance [105].

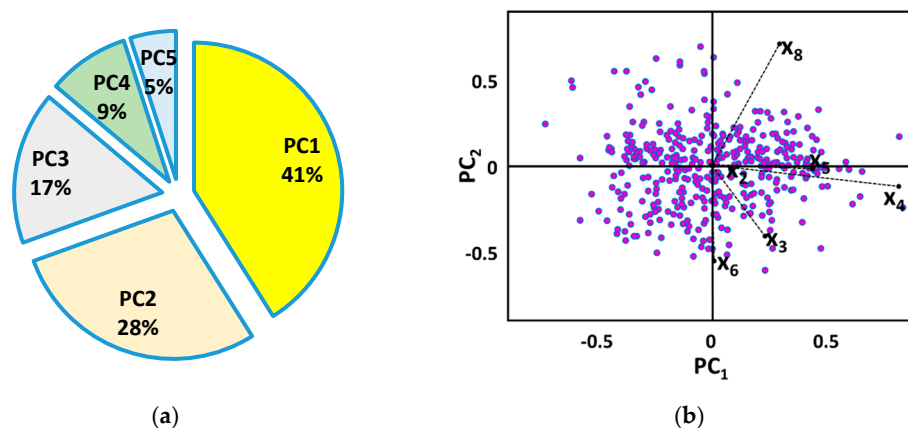


Figure 4. (a) Variance in the five most influential principal components and (b) the relationship between the two variables and PC2.

The results from four regression models (linear, 2FI, quadratic, and reduced quadratic) are presented in Table 4 to present the error rates for the training and testing stages. These error rates helped to inform our determinations of the model efficiency after a thorough evaluation of all 50 training and testing datasets produced using the k -fold method. From these data, it is apparent that the 2FI regression model was best suited to all nine independent variables. The quadratic model performed best with the principal components (PC₁ to PC₅) and represented the lowest mean and standard deviation of the RMSE, MAPE, and EF. Limiting the models to the principal components PC₁ to PC₅ did not improve the MLR model yield prediction. The model accuracy decreased in the testing rather than the training phase (Table 3). Therefore, MLR models have an acceptable level of generalizability. Emamgholizadeh et al. (2015) applied an MLR model in an investigation of the agronomic traits and yield of sesame [16]. The ML-PCA and ML-no-PCA exhibited approximately the same performance in terms of the prediction accuracy based on the RMSE. Comparatively, the MAPE showed a maximum ~3% higher accuracy using ML-no-PCA in comparison with ML-PCA. However, ML-PCA is fast and has lower computing expenses compared with ML-no-PCA.

Table 4. Comparison of the yield predictions of four MLR model types before and after PCA exclusion.

Phase		Train			Test		
Model		RMSE	MAPE	EF	RMSE	MAPE	EF
Linear	no-PCA	0.29 ± 0.01	8.04 ± 0.80	0.82 ± 0.01	0.29 ± 0.03	8.36 ± 0.80	0.80 ± 0.04
	PCA	0.36 ± 0.01	11.21 ± 0.26	0.72 ± 0.01	0.37 ± 0.02	11.43 ± 1.02	0.71 ± 0.05
2FI	no-PCA	0.21 ± 0.01	5.82 ± 0.15	0.90 ± 0.01	0.25 ± 0.02	7.02 ± 0.68	0.86 ± 0.02
	PCA	0.35 ± 0.01	10.27 ± 0.32	0.74 ± 0.02	0.38 ± 0.04	11.11 ± 1.14	0.67 ± 0.10
Quadratic	no-PCA	0.21 ± 0.01	5.68 ± 0.15	0.90 ± 0.01	0.26 ± 0.02	7.15 ± 0.67	0.84 ± 0.33
	PCA	0.31 ± 0.00	8.97 ± 0.26	0.80 ± 0.01	0.34 ± 0.03	9.89 ± 1.03	0.75 ± 0.05
Reduced quadratic	no-PCA	0.25 ± 0.01	6.52 ± 0.22	0.86 ± 0.01	0.26 ± 0.03	6.96 ± 0.80	0.84 ± 0.03
	PCA	0.32 ± 0.01	9.37 ± 0.24	0.78 ± 0.01	0.34 ± 0.03	8.87 ± 0.80	0.76 ± 0.03

Mokarram and Bijanzadeh (2016) predicted barley (*Hordeum vulgare*) grain yields using the percentage of soil organic matter and grain/spike ratio as independent variables for MLR model inputs. They observed that the machine learning model ANN ($R^2 = 0.922$) performed more accurately than MLR ($R^2 = 0.784$) [85]. The four transformation types shown in Table 5 were assessed for their ability to improve the MLR model performance. A comparison between the achieved results (Tables 4 and 5) revealed that the utilization of response variable transformation (y) did not enhance the model prediction efficiency. Thus, the final MLR model was created without transformation. The k -fold method generated 50 different datasets from among which the “optimal” dataset was selected for the assessment of the model training and validation steps. The results gained after subjecting this dataset to ANOVA with 2FI are presented in Table 6. Correlation tests between the independent variables and yield showed an insignificant correlation between the yield and the number of fruit-bearing branches (x_9), leading to its exclusion from the model. A p -value cutoff of 0.10 was employed to aid in the selection of the final model and effective traits. This stringency level resulted in 11 traits (x_2 , x_3 , x_6 , x_8 , x_{12} , x_{16} , x_{18} , x_{37} , x_{67} , x_{68} , and x_{78}) selected from a total of 36 using a stepwise regression process. Among the main variables, the days to 100% flowering (x_2), days to maturity (x_3), CPP (x_6), and SPC (x_8) were directly incorporated into the model. Other variables were included after viewing the interactions between the main independent variables. El-Mohsen (2013) discovered a relationship between the yield and agro-morphological traits through a stepwise regression process, which showed that 77.25% of the variance in the SSY was explained by the days to flowering and CPP [27]. Additionally, Parimala and Mathur (2006) indicated that the CPP was the most effective factor for predicting the yield [25]. Yol et al. (2010) defined selection inputs based on the pH, CPP, BN, and TSW to elucidate the SSY [106].

Table 5. MLR model performance parameters for four different transformation types.

Phase	Train			Test		
TV *	EF	MAPE	RMSE	EF	MAPE	RMSE
$\sqrt{y}$	0.90 ± 0.00	5.99 ± 0.17	0.22 ± 0.00	0.84 ± 0.03	7.32 ± 0.70	0.27 ± 0.02
$\ln y$	0.88 ± 0.00	6.34 ± 0.26	0.23 ± 0.00	0.79 ± 0.05	7.86 ± 0.81	0.31 ± 0.04
$1/y$	0.06 ± 3.09	7.87 ± 0.79	0.49 ± 0.46	-0.85 ± 5.80	11.11 ± 4.40	0.92 ± 1.42
y^2	0.84 ± 0.01	6.65 ± 0.12	0.27 ± 0.01	0.75 ± 0.20	9.58 ± 0.91	0.28 ± 0.32

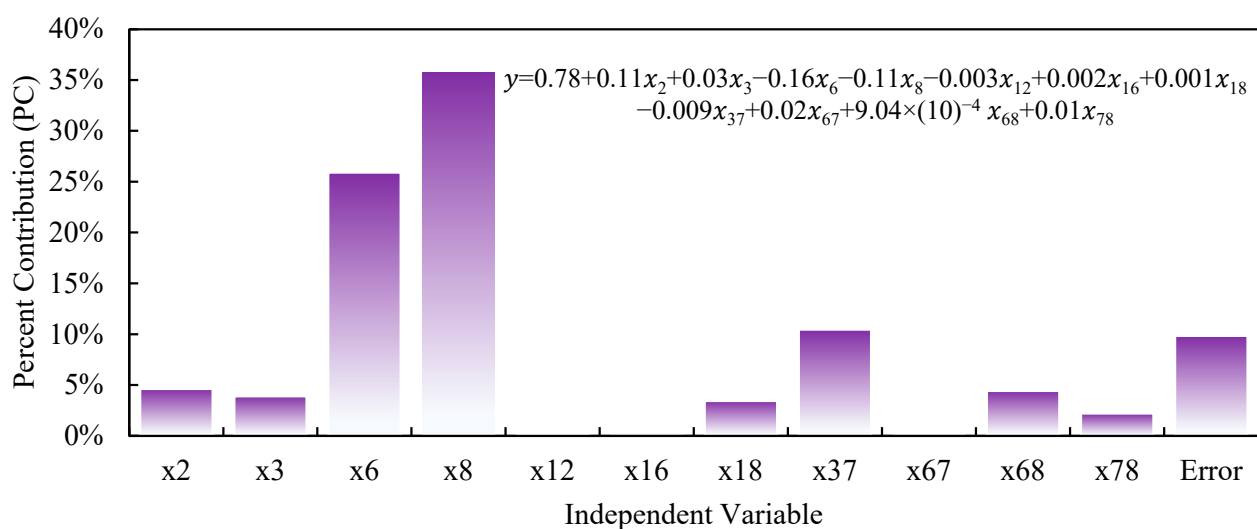
* TV, transformation variable.

Table 6. MLR regression model ANOVA for the optimal training and validation dataset.

Source	DF	SS	p	Source	DF	SS	p	Source	DF	SS	p
Model	36	135	0	FT10*PP	1	0.34	0.1	SM*SPC	1	1.03	0.9
FT10	1	7.3	0.48	FT10*TSW	1	0	0.65	PH*PHN	1	0.01	0.47
FT100	1	0.55	0.07	FT10*SPC	1	0.56	0.1	PH*CPP	1	0.27	0.02
SM	1	5.19	0.09	FT100*SM	1	0	0.8	PH*TSW	1	0.04	0.93
PH	1	19.52	0.54	FT100*PH	1	0	0.79	PH*SPC	1	0.27	0.66
PHN	1	0.23	0.69	FT100*PHN	1	0	0.42	PHN*CPP	1	1.08	0.51
CPP	1	36.36	0	FT100*CPP	1	0.04	0.63	PHN*TSW	1	0.3	0.18
TSW	1	1.9	0.5	FT100*TSW	1	0	0.68	PHN*SPC	1	0	0.73
SPC	1	51.08	0.02	FT100*SPC	1	0.35	0.11	CPP*TSW	1	0	0
FT10*100	1	1.05	0.01	SM*PH	1	0	0.49	CPP*TSW	1	3.95	0
FT10*SM	1	0.17	0.32	SM*PHN	1	0.02	0.27	TSW*SPC	1	1.43	0
FT10*PH	1	0.31	0.37	SM*CPP	1	0.1	0.9	Residual	265	14.61	-
FT10*PHN	1	0.03	0.51	SM*TSW	1	1.17	0.05	Total	301	149.3	-

* Interaction function between the two independent variables; p-value (p); sum of squares (SS).

The percentage contribution (PC) for the features applied for the estimations of the SSY is generated by dividing the sum of the squares of each feature by the total sum of the squares. Presented in Figure 5, the PC of error is 9.84% in the training step, with the CPP (x_6) and SPC (x_8) displaying the highest PC. As interaction factors such as x_{12} (FT10*FT100), x_{16} (FT10*CPP), and x_{67} (CPP*TSW) have the lowest PC, it can be assumed that the x_6 and x_8 variables are very effective parameters for SSY estimation. The FT10*SPC (x_{18}), SM*TSW (x_{37}), CPP*SPC (x_{68}), and TSW*SPC (x_{78}) interaction variables also contributed significantly to determining the SSY.

**Figure 5.** Percent contribution (PC) for each MLR model factor and PC of error in the training phase.

3.3. ML Models Evaluation

The evaluation of the model variables with and without PCA allowed for our effective observation of the input accuracy in predicting the SSY. In addition to identifying specific input sets which can influence the effectiveness of the ML model (MLR, GPR, and RBF) prediction, tuning the hidden size number is crucial for RBF. Determining the hidden layer neuron numbers (i.e., hidden layer size) is a key first step in reliable neural network model design. The trial and error process enables the adjustment of the neuron numbers by providing data, similar to factorial analysis, for the optimization of the hidden layer size (Table 6). The results of the RMSE, MAPE, and EF in the training and test steps using the 50 unique datasets indicated that the highest predicted performance of the RBF resulted in 20 neurons in the hidden layer. The complete results of the MLR, GPR, and RBF models during the training, test, and combined phases are presented in Table 7. These data are displayed without PCA, using the two classes of the original factors (see Equation (18)), and with PCA, incorporating principal components PC₁ to PC₅. The addition of the PCA-based variables did not enhance the MLR, GPR, and RBF model performance. The dataset performance, in Table 8, was gauged by considering the lowest RMSE, TSSE, and MAPE values and the highest EF value.

Table 7. RBF neural network performance factorial analysis with varied hidden layer neuron numbers.

Phase	Criteria	Hidden Layer Size							
		5	10	15	20	25	30	35	40
Train	RMSE	0.22	0.21	0.21	0.2	0.2	0.2	0.2	0.2
	MAPE	6.31	5.99	5.85	5.74	5.81	5.75	5.8	5.81
	EF	0.89	0.9	0.9	0.91	0.91	0.91	0.91	0.91
Test	RMSE	0.26	0.25	0.26	0.26	0.26	0.26	0.26	0.26
	MAPE	7.31	7.09	7.24	7.15	7.14	7.17	7.16	7.21
	EF	0.86	0.85	0.86	0.85	0.85	0.85	0.85	0.85
Total	RMSE	0.23	0.22	0.22	0.22	0.22	0.22	0.22	0.22
	MAPE	6.51	6.21	6.13	6.02	6.08	6.04	6.07	6.09
	EF	0.89	0.89	0.89	0.9	0.9	0.9	0.9	0.9

Table 8. MLR, GPR, and RBF model assessments based on initial and PCA-derived input variables.

-PCA *		Train			Test			Total		
Model	MLR	GPR	RBF	MLR	GPR	RBF	MLR	GPR	RBF	
RMSE	0.23	0.05 ± 0.06	0.20 ± 0.01	0.23	0.26 ± 0.03	0.26 ± 0.02	0.23	0.13 ± 0.03	0.22 ± 0.01	
TSSE	16.25	2.17 ± 3.48	13.10 ± 1.74	3.90	5.46 ± 1.27	5.27 ± 1.05	20.15	7.63 ± 3.45	18.38 ± 1.30	
MAPE	6.29	1.37 ± 1.83	5.74 ± 0.38	5.89	7.11 ± 0.71	7.15 ± 0.76	6.21	2.52 ± 1.44	6.02 ± 0.25	
EF	0.89	0.98 ± 0.02	0.91 ± 0.01	0.90	0.85 ± 0.03	0.85 ± 0.03	0.89	0.95 ± 0.01	0.90 ± 0.01	
+PCA *		Train			Test			Total		
Model	MLR	GPR	RBF	MLR	GPR	RBF	MLR	GPR	RBF	
RMSE	0.36	0.21 ± 0.09	0.30 ± 0.01	0.32	0.34 ± 0.02	0.30 ± 0.01	0.35	0.25 ± 0.04	0.30 ± 0.01	
TSSE	38.92	16.89 ± 8.77	27.88 ± 2.95	7.77	9.16 ± 1.49	27.88 ± 2.95	46.69	26.06 ± 8.17	27.88 ± 2.95	
MAPE	11.01	6.31 ± 2.84	8.89 ± 0.42	8.55	9.97 ± 0.91	8.89 ± 0.42	10.52	7.05 ± 2.22	8.89 ± 0.42	
EF	0.74	0.88 ± 0.05	0.81 ± 0.01	0.80	0.74 ± 0.03	0.81 ± 0.01	0.75	0.86 ± 0.04	0.81 ± 0.01	

* -, excluding PCA (using nine original input variables); +, including PCA (using five principal components (PCs) extracted from nine variables).

The results of the GPR and RBF model performance in Table 9 were more accurate than the GPR-PCA and RBF-PCA in both the training and testing steps. Generally, the GPR, RBF and MLR showed the best prediction performance in the training phase (TSSE

of 16.89, 27.88, and 38.92, respectively). The MLR was the ideal model in the test phase, with a TSSE of 7.77. The RMSE displayed non-significant sensitivity in the ML comparison. Since the training step performance of the GRP is reliable, no overfitting problems are observed. Although the previous results (Table 8) are acceptable, hidden-layer optimization procedures improve the performance accuracy of the GPR and RBF models. Additionally, as reported for the previous models, the RBF model prediction performance is improved with the use of the most influential variables compared with principal components PC₁ to PC₅ (RBF-PCA) as neural network inputs.

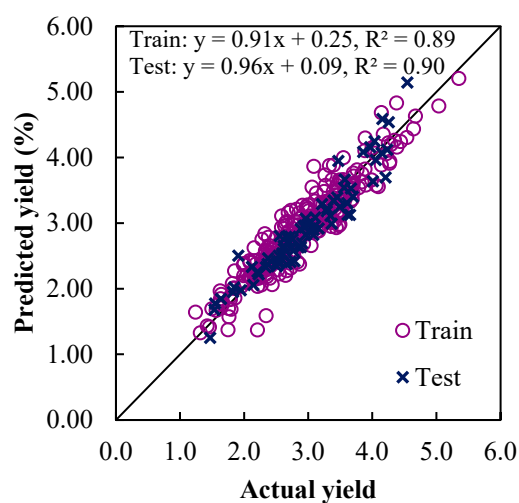
Table 9. RBF and GPR model performance with the optimal hidden layer neuron number.

Phase	Train				Test				Total			
Model	RBF	RBF-PCA	GPR	GPR-PCA	RBF	RBF-PCA	GPR	GPR-PCA	RBF	RBF-PCA	GPR	GPR-PCA
RMSE	0.21	0.30	0.00	0.29	0.23	0.36	0.21	0.30	0.21	0.31	0.09	0.29
TSSE	12.75	27.17	0.00	26.01	4.08	9.77	3.22	6.75	16.83	36.94	3.22	32.76
MAPE	5.72	8.75	0.02	8.60	6.39	9.89	5.48	9.09	5.86	8.98	1.12	8.70
EF	0.91	0.81	0.99	0.83	0.91	0.78	0.90	0.81	0.91	0.80	0.98	0.82

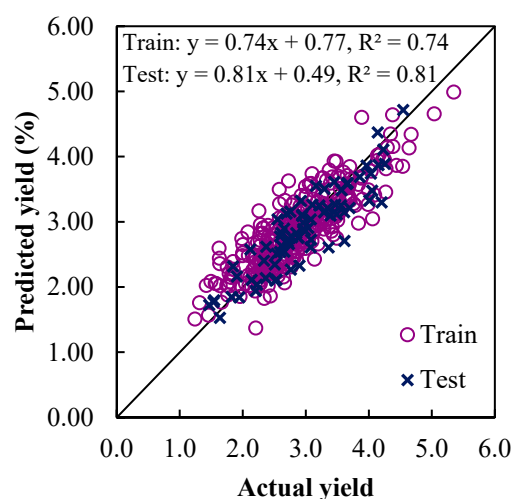
3.4. Prediction of Sesame-Seed-Yield-Based ML Models

The distribution of the predicted and actual SSY values in Figure 6 between the ML and ML-PCA models reveals similar performances between the training and testing steps. The fit results for the MLR and MLR-PCA models are illustrated in Figure 6a,b, with an R^2 value of 0.89 and 0.90 in the training and testing steps, respectively. The distribution patterns of the actual and predicted data in Figure 6c,d for the GPR and GPR-PCA models were comparable to those of the MLR for the test phase ($R^2 = 0.89$), but the value for the test phase was ten percent higher ($R^2 = 0.99$) than in the training phase. An R^2 value of 0.91, presented in Figure 6e,f, was identified for the training and test phases of the RFP model. The ML-PCA models all exhibited lower R^2 values in the training and testing phases for the prediction of the SSY, an indication that the ML models were less precise when combined with the PCA technique. Based on the R^2 values reported here, the use of agro-morphological features is an acceptable method when applying models to predict the SSY. The frequency distributions in Figure 7 aid in the evaluations of the ML and ML-PCA model performance across several error ranges.

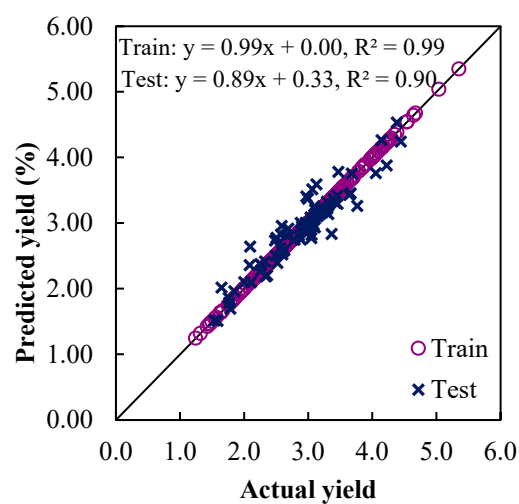
The MLR model error ranged from -0.24 to 0.84 , with 78% of the error falling between -0.24 and 0.30 . The error for the MLR-PCA model ranged from -0.30 to 1.02 , with 60% of the error between -0.30 and 0.36 . The error distribution for the GPR and GPR-PCA models implies that the GPR model is superior to the GPR-PCA model. Approximately 94% of the errors in the GPR model are between -0.19 and 0.17 , while the range for all the errors is from -0.19 to 0.53 . Additionally, the error distributions of the RBF and RBF-PCA models are provided in three intervals, similar to the other MLs mentioned above. Approximately 71% of the errors in the RBF model fall between -0.15 and 0.26 , while 74% of the errors fall between -0.21 and 0.54 in the RBF-PCA model.



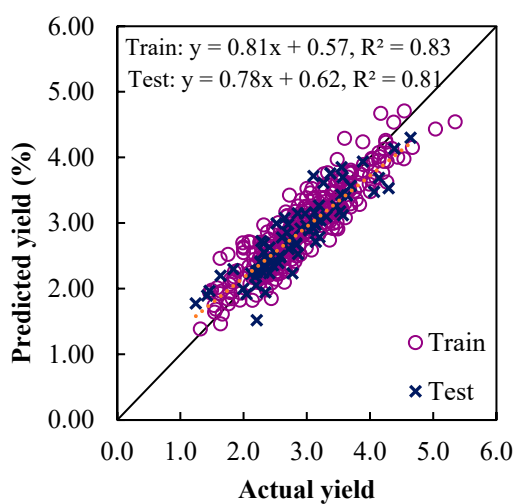
(a) MLR



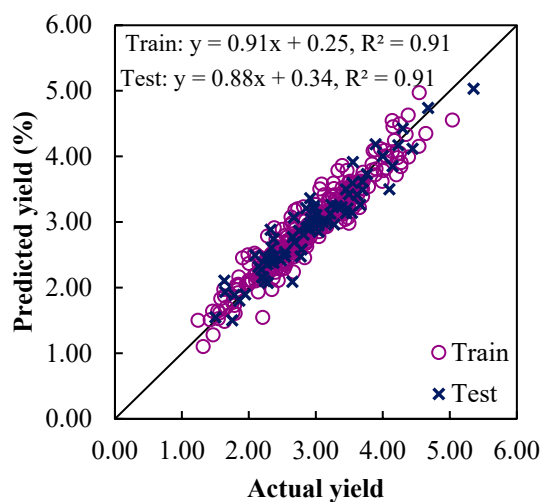
(b) MLR-PCA



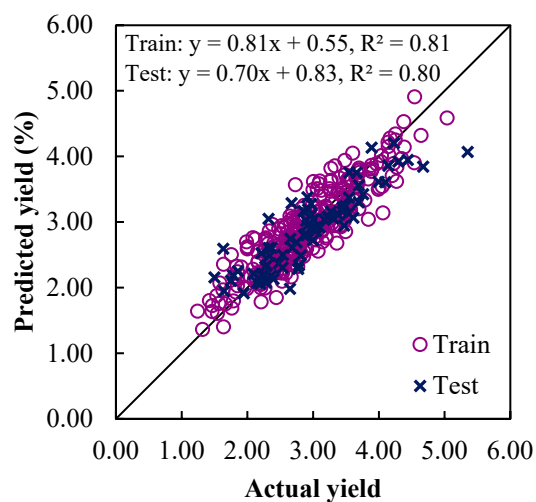
(c) GPR



(d) GPR-PCA



(e) RBF



(f) RBF-PCA

Figure 6. Correlations between actual and predicted SSY values of three ML and ML-PCA models.

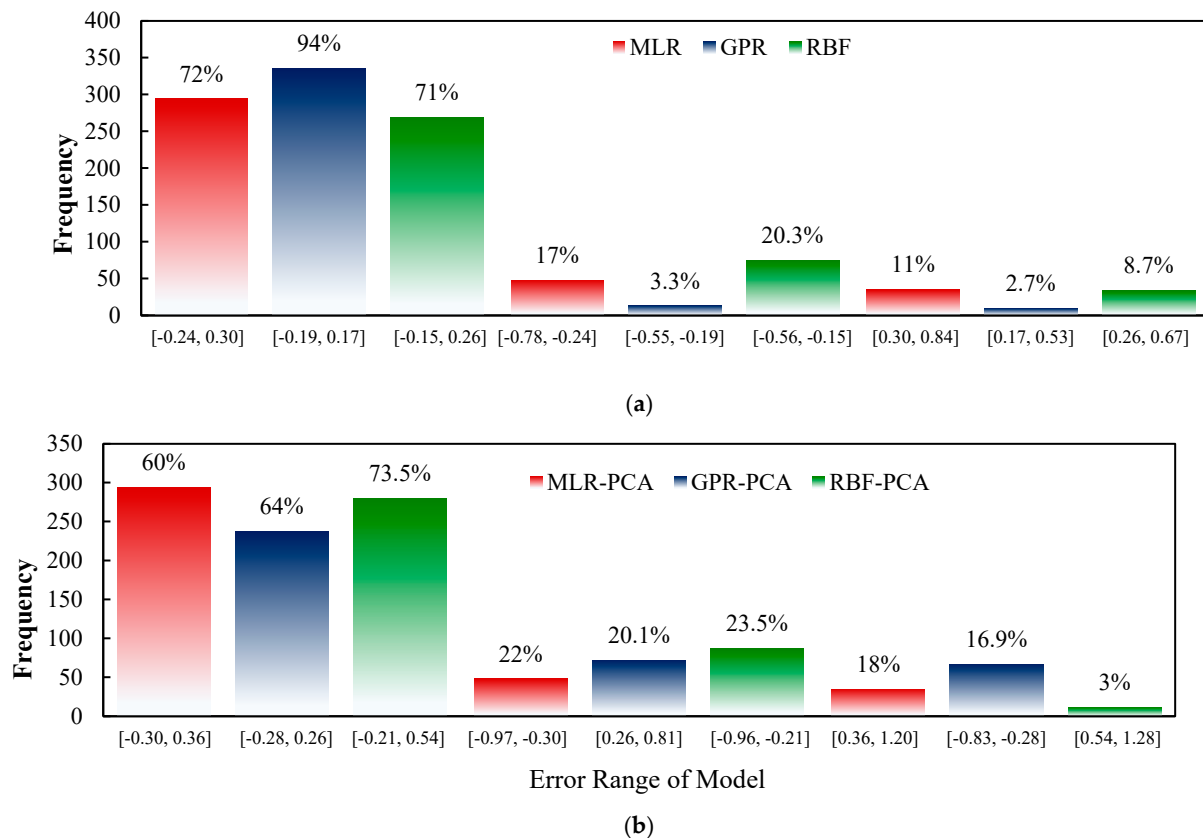


Figure 7. Frequency distribution of MLR, GPR, and RBF errors for (a) initial and (b) PCA independent variables.

3.5. Sensitivity Analysis

Uncertainty, in ML modeling, is best visualized through a sensitivity analysis. Conventionally categorized into one of two classes, the global or local, sensitivity analyses determine how data results change in response to alterations in ML models or methods [107,108]. A local sensitivity analysis only surveys a specific area while leaving the remaining inputs or independent variables unexplored, whereas global sensitivity analyses simultaneously incorporate the variance in all independent variables to ascertain a nonlinear trend across a range of variables [109]. These tests make it easy to distinguish the comparative importance of each input variable for the SSY. As illustrated in Figure 8, the GPR, MLR, and RBF model sensitivity indices are indicative of the model performance after the exclusion of the agro-morphological variables as inputs. Measurements of the sensitivity analysis data were collected and presented with the RMSE, MAPE, and EF. Each factor (x_1 to x_9) was individually excluded to isolate the level of sensitivity for each variable. Manipulations of the CPP (x_6), SPC (x_8), and TSW (x_7) variables were the most influential on the SSY prediction precision, with error rates between 0.29 and 0.55 for the RMSE, as shown in Figure 8a.

The most sensitive ML models for each of these variables were the MLR, RBF, and GPR, respectively. When these ML models were assessed based on the MAPE index, as shown in Figure 8b, this revealed similar results. The highest MAPEs, those from the CPP (14.2–11.06%), SPC (13.81–7.18%), and TSW (8.43–6.37%), were recorded after the MLR, RBF, and GPR model exclusion. The results of the EF in Figure 8c exposed a decreasing trend when the CPP, SPC, and TSW were removed from the model input. Notable errors were revealed in the case of other agro-morphological indicators, providing further evidence of the necessity for careful variable selection. The obtained model errors illustrate the fact that their exclusion reduces the efficiency of the ML models, including the MLR, GPR, and RBF, by 50%.

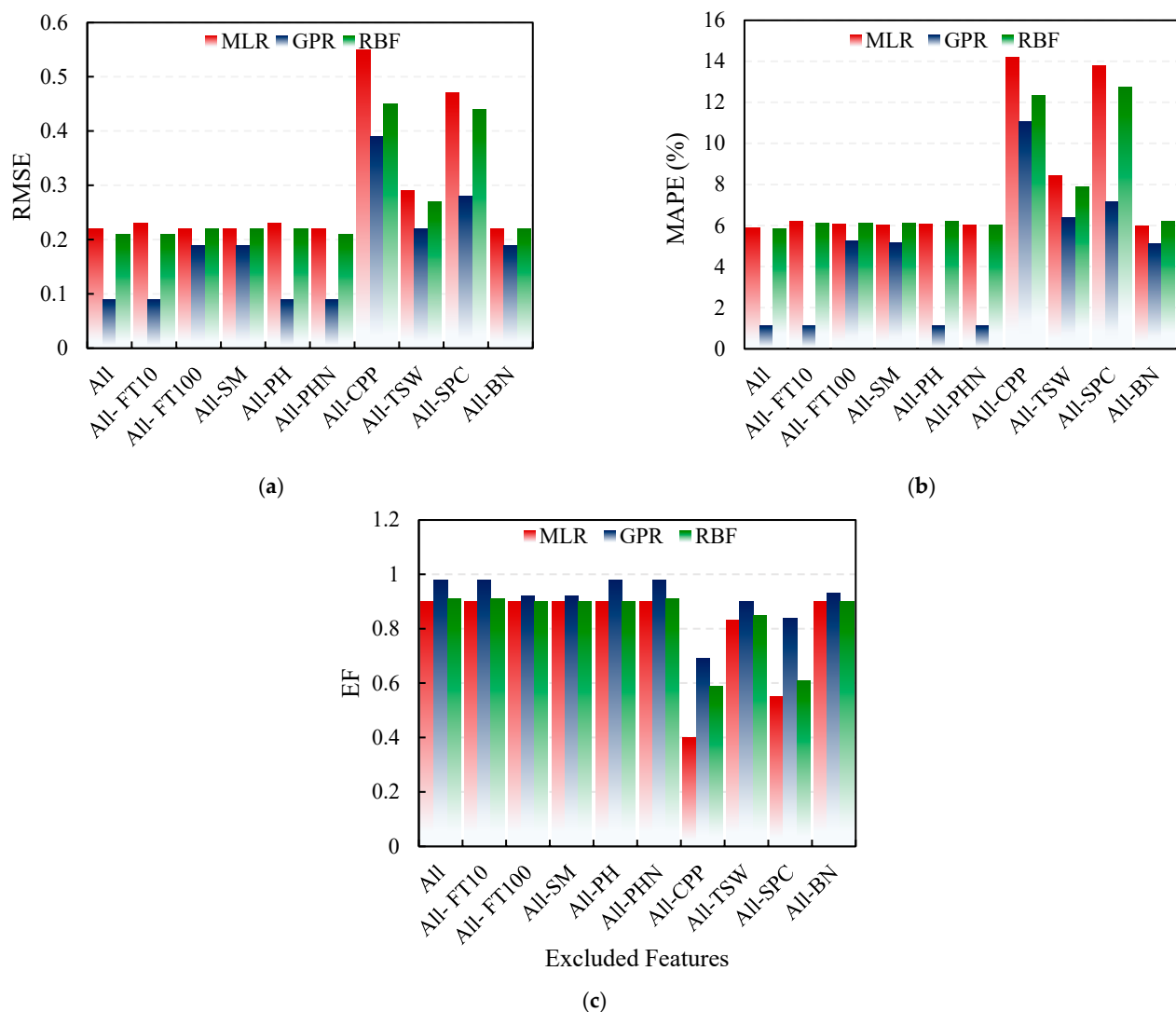


Figure 8. Sensitivity analysis results for MLR, GPR, and RBF models to assess the SSY based on the (a) RMSE, (b) MAPE and (c) EF criteria.

4. Conclusions

Here, multi-machine learning (ML) approaches were applied to generate predictions of the seed yield in sesame. The initial study began with the MLR model and PCA analysis, combined with data from existing studies. Nine agro-morphological factors contributed to the establishment of multi-ML techniques, including the MLR, GPR, and RBF, used for predictions of the SSY. Moreover, the coupled PCA-ML models incorporated five principal components from nine primary variables and, for prediction accuracy, were compared with the original ML models. The results obtained here suggest that the SSY can be effectively estimated by the GPR, RBF, and MLR models. The SSY estimated by the GPR, RBF, and MLR had RMSEs of 0 to 0.05 t/ha, 0.20 to 0.23 t/ha, and 0.23 to 0.36 t/ha, respectively. The best performance was displayed by the MLR model, subjected to the 2FI inputs. The RMSE declined in the case of the GPR and RBF models after the k-fold method was employed to assist in the dataset selection for the training and validation steps. The GPR model predicted SSY more accurately and precisely than MLR or RBF. The use of global sensitivity analysis indicated that the CPP, SPC, and TSW were the most sensitive factors for the sesame yield estimation. These findings could be vital in efforts to promote productivity by planting more robust sesame species. The agro-morphological characteristics investigated here can be coupled with phenolic and spectral data analyses in future studies to implement a multi-

objective modeling approach. Additional studies of traits such as the yield and quality level could provide the data necessary to formulate a solid and sustainable plan in order to overcome several primary challenges in oilseed cropping systems. In future research, the physio-chemical inputs, such as chemical components, could promote prediction at a higher level. Additionally, the ML can present the optimal geographical location for the plant seeds necessary to obtain higher yields by merging climate input variables and convolutional neural network (CNN) models.

Author Contributions: M.P.: Writing-Original draft preparation, Data acquisition and Analysis. M.R.: Data curation, Editing draft preparation and Analysis. A.R.: Participation in the concept, Software, Validation, Writing-Original draft preparation. S.S.L.: Editing, Writing and revision. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: All participants consented to the processing of their data in anonymous and confidential form. Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Acknowledgments: The authors acknowledge and appreciate the funding and technical support provided by the Shahrood University of Technology and Ferdowsi University of Mashhad, Iran, for this research.

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

References

1. Bassegio, D.; Zanutto, M.D.; Santos, R.F.; Werncke, I.; Dias, P.P.; Olivo, M. Oilseed crop crambe as a source of renewable energy in Brazil. *Renew. Sustain. Energy Rev.* **2016**, *66*, 311–321. [\[CrossRef\]](#)
2. Mousavi-Avval, S.H.; Shah, A. Techno-economic analysis of hydroprocessed renewable jet fuel production from pennycress oilseed. *Renew. Sustain. Energy Rev.* **2021**, *149*, 111340. [\[CrossRef\]](#)
3. Ikegami, M.; Wang, Z. Does energy aid reduce CO₂ emission intensities in developing countries? *J. Environ. Econ. Policy* **2021**, *10*, 343–358. [\[CrossRef\]](#)
4. Agidew, M.G.; Dubale, A.A.; Atlabachew, M.; Abebe, W. Fatty acid composition, total phenolic contents and antioxidant activity of white and black sesame seed varieties from different localities of Ethiopia. *Chem. Biol. Technol. Agric.* **2021**, *8*, 14. [\[CrossRef\]](#)
5. Muthulakshmi, C.; Sivaranjani, R.; Selvi, S. Modification of sesame (*Sesamum indicum* L.) for Triacylglycerol accumulation in plant biomass for biofuel applications. *Biotechnol. Rep.* **2021**, *32*, e00668. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Benami, E.; Jin, Z.; Carter, M.R.; Ghosh, A.; Hijmans, R.J.; Hobbs, A.; Kenduiywo, B.; Lobell, D.B. Uniting remote sensing, crop modelling and economics for agricultural risk management. *Nat. Rev. Earth Environ.* **2021**, *2*, 140–159. [\[CrossRef\]](#)
7. Wang, Y.; Li, X.; Lee, T.; Peng, S.; Dou, F. Effects of nitrogen management on the ratoon crop yield and head rice yield in South USA. *J. Integr. Agric.* **2021**, *20*, 1457–1464. [\[CrossRef\]](#)
8. Hiremath, S.C.; Patil, C.G.; Patil, K.B.; Nagasampige, M.H. Genetic diversity of seed lipid content and fatty acid composition in some species of *Sesamum* L. (Pedaliaceae). *Afr. J. Biotechnol.* **2007**, *6*, 539–543.
9. Uzun, B.; Arslan, Ç.; Furat, Ş. Variation in fatty acid compositions, oil content and oil yield in a germplasm collection of sesame (*Sesamum indicum* L.). *J. Am. Oil Chem. Soc.* **2008**, *85*, 1135–1142. [\[CrossRef\]](#)
10. Han, L.; Li, J.; Wang, S.; Cheng, W.; Ma, L.; Liu, G.; Han, D.; Niu, L. Sesame oil inhibits the formation of glycidyl ester during deodorization. *Int. J. Food Prop.* **2021**, *24*, 505–516. [\[CrossRef\]](#)
11. Karrar, E.; Ahmed, I.A.M.; Manzoor, M.F.; Al-Farga, A.; Wei, W.; Albakry, Z.; Sarpong, F.; Wang, X. Effect of roasting pretreatment on fatty acids, oxidative stability, tocopherols, and antioxidant activity of gurum seeds oil. *Biocatal. Agric. Biotechnol.* **2021**, *34*, 102022. [\[CrossRef\]](#)
12. Mahmood, T.; Mustafa HSBin Aftab, M.; Ali, Q.; Malik, A. Super canola: Newly developed high yielding, lodging and drought tolerant double zero cultivar of rapeseed (*Brassica napus* L.). *Genet. Mol. Res.* **2019**, *18*, gmr16039951.
13. Tadesse, T.; Singh, H.; Weyessa, B. Correlation and path coefficient analysis among seed yield traits and oil content in Ethiopian linseed germplasm. *Int. J. Sustain. Crop Prod.* **2009**, *4*, 8–16.
14. Solanki, Z.S.; Gupta, D. Inheritance studies for seed yield in sesame. *Sesame Safflower Newsl.* **2003**, *18*, 25–28.
15. Khan, M.A.; Mirza, M.Y.; Akmal, M.; Ali, N.; Khan, I. Genetic parameters and their implications for yield improvement in sesame. *Sarhad J. Agric.* **2007**, *23*, 623.

16. Emamgholizadeh, S.; Parsaeian, M.; Baradaran, M. Seed yield prediction of sesame using artificial neural network. *Eur. J. Agron.* **2015**, *68*, 89–96. [\[CrossRef\]](#)
17. Shastry, A.; Sanjay, H.A.; Bhanusree, E. Prediction of crop yield using regression techniques. *Int. J. Soft Comput.* **2017**, *12*, 96–102.
18. Sellam, V.; Poovammal, E. Prediction of crop yield using regression analysis. *Indian J. Sci. Technol.* **2016**, *9*, 1–5. [\[CrossRef\]](#)
19. Ramesh, D.; Vardhan, B.V. Analysis of crop yield prediction using data mining techniques. *Int. J. Res. Eng. Technol.* **2015**, *4*, 47–473.
20. Chowdhury, S.; Datta, A.K.; Saha, A.; Sengupta, S.; Paul, R.; Maity, S.; Das, A. Traits influencing yield in sesame (*Sesamum indicum* L.) and multilocal trials of yield parameters in some desirable plant types. *Indian J. Sci. Technol.* **2010**, *3*, 163–166. [\[CrossRef\]](#)
21. Sengupta, S.; Datta, A.K. Genetic studies to ascertain selection criteria for yield improvement in sesame. *J. Phytol. Res.* **2004**, *17*, 163–166.
22. Shim, K.B.; Kang, C.W.; Lee, S.W.; Kim, D.H.; Lee, B.H. Heritabilities, genetic correlations and path coefficients of some agronomic traits in different cultural environments in sesame. *Sesame Safflower Newsl.* **2001**, 16–22.
23. Boureima, S.; Diouf, S.; Amoukou, M.; Van Damme, P. Screening for sources of tolerance to drought in sesame induced mutants: Assessment of indirect selection criteria for seed yield. *Int. J. Pure Appl. Biosci.* **2016**, *4*, 45–60. [\[CrossRef\]](#)
24. Ganesh, S.K.; Sakila, M. Association analysis of single plant yield and its yield contributing characters in sesame (*Sesamum indicum* L.). *Sesame Safflower Newsl.* **1999**, *14*, 16–19.
25. Parimala, K.; Mathur, R.K. Yield component analysis through multiple regression analysis in sesame. *Int. J. Agric. Res.* **2006**, *2*, 338–340.
26. Shim, K.-B.; Kang, C.-W.; Seong, J.-D.; Hwang, C.-D.; Suh, D.-Y. Interpretation of relationship between sesame yield and its components under early sowing cropping condition. *Korean J. Crop Sci.* **2006**, *51*, 269–273.
27. Abd El-Mohsen, A.A. Comparison of some statistical techniques in evaluating Sesame yield and its contributing factors. *Scientia* **2013**, *1*, 8–14.
28. Soltanali, H.; Rohani, A.; Tabasizadeh, M.; Abbaspour-Fard, M.H.; Parida, A. An improved fuzzy inference system-based risk analysis approach with application to automotive production line. *Neural. Comput. Appl.* **2020**, *32*, 10573–10591. [\[CrossRef\]](#)
29. Shin, M.; Ithnin, M.; Vu, W.T.; Kamaruddin, K.; Chin, T.N.; Yaakub, Z.; Chang, P.L.; Sritharan, K.; Nuzhdin, S.; Singh, R. Association mapping analysis of oil palm interspecific hybrid populations and predicting phenotypic values via machine learning algorithms. *Plant Breed* **2021**, *140*, 1150–1165. [\[CrossRef\]](#)
30. Wen, G.; Ma, B.-L.; Vanasse, A.; Caldwell, C.D.; Earl, H.J.; Smith, D.L. Machine learning-based canola yield prediction for site-specific nitrogen recommendations. *Nutr. Cycl. Agroecosyst.* **2021**, *121*, 241–256. [\[CrossRef\]](#)
31. Sharma, R.; Kamble, S.S.; Gunasekaran, A.; Kumar, V.; Kumar, A. A systematic literature review on machine learning applications for sustainable agriculture supply chain performance. *Comput. Oper. Res.* **2020**, *119*, 104926. [\[CrossRef\]](#)
32. Rahimi, M.; Abbaspour-Fard, M.H.; Rohani, A. A multi-data-driven procedure towards a comprehensive understanding of the activated carbon electrodes performance (using for supercapacitor) employing ANN technique. *Renew. Energy* **2021**, *180*, 980–992. [\[CrossRef\]](#)
33. Xu, H.; Zhang, X.; Ye, Z.; Jiang, L.; Qiu, X.; Tian, Y.; Zhu, Y.; Cao, W. Machine learning approaches can reduce environmental data requirements for regional yield potential simulation. *Eur. J. Agron.* **2021**, *129*, 126335. [\[CrossRef\]](#)
34. Wang, X.; Miao, Y.; Dong, R.; Zha, H.; Xia, T.; Chen, Z.; Kusnierek, K.; Mi, G.; Sun, H.; Li, M. Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn. *Eur. J. Agron.* **2021**, *123*, 126193. [\[CrossRef\]](#)
35. Rahimi, M.; Abbaspour-Fard, M.H.; Rohani, A.; Yuksel Orhan, O.; Li, X. Modeling and Optimizing N/O-Enriched Bio-Derived Adsorbents for CO₂ Capture: Machine Learning and DFT Calculation Approaches. *Ind. Eng. Chem. Res.* **2022**, *61*, 10670–10688. [\[CrossRef\]](#)
36. Soltanali, H.; Nikkhah, A.; Rohani, A. Energy audit of Iranian kiwifruit production using intelligent systems. *Energy* **2017**, *139*, 646–654. [\[CrossRef\]](#)
37. Nikkhah, A.; Rohani, A.; Rosentrater, K.A.; El Haj Assad, M.; Ghnimi, S. Integration of principal component analysis and artificial neural networks to more effectively predict agricultural energy flows. *Environ. Prog. Sustain. Energy* **2019**, *38*, 13130. [\[CrossRef\]](#)
38. Taki, M.; Mehdizadeh, S.A.; Rohani, A.; Rahnama, M.; Rahmati-Joneidabad, M. Applied machine learning in greenhouse simulation; new application and analysis. *Inf. Process. Agric.* **2018**, *5*, 253–268. [\[CrossRef\]](#)
39. Bolandnazar, E.; Rohani, A.; Taki, M. Energy consumption forecasting in agriculture by artificial intelligence and mathematical models. *Energy Sources Part A Recover. Util. Environ. Eff.* **2020**, *42*, 1618–1632. [\[CrossRef\]](#)
40. Rahimi, M.; Abbaspour-Fard, M.H.; Rohani, A. Synergetic effect of N/O functional groups and microstructures of activated carbon on supercapacitor performance by machine learning. *J. Power Source* **2022**, *521*, 230968. [\[CrossRef\]](#)
41. Jayas, D.S.; Paliwal, J.; Visen, N.S. Review paper (AE—Automation and emerging technologies): Multi-layer neural networks for image analysis of agricultural products. *J. Agric. Eng. Res.* **2000**, *77*, 119–128. [\[CrossRef\]](#)
42. Karimi, Y.; Prasher, S.O.; McNairn, H.; Bonnell, R.B.; Dutilleul, P.; Goel, P.K. Classification accuracy of discriminant analysis, artificial neural networks, and decision trees for weed and nitrogen stress detection in corn. *Trans. ASAE* **2005**, *48*, 1261–1268. [\[CrossRef\]](#)
43. Elizondo, D.; Hoogenboom, G.; McClendon, R.W. Development of a neural network model to predict daily solar radiation. *Agric. For. Meteorol.* **1994**, *71*, 115–132. [\[CrossRef\]](#)

44. Mukerji, A.; Chatterjee, C.; Raghuwanshi, N.S. Flood forecasting using ANN, neuro-fuzzy, and neuro-GA models. *J. Hydrol. Eng.* **2009**, *14*, 647–652. [\[CrossRef\]](#)
45. Rahimi, M.; Abbaspour-Fard, M.H.; Rohani, A. Machine learning approaches to rediscovery and optimization of hydrogen storage on porous bio-derived carbon. *J. Clean. Prod.* **2021**, *329*, 129714. [\[CrossRef\]](#)
46. Jin, Y.-Q.; Liu, C. Biomass retrieval from high-dimensional active/passive remote sensing data by using artificial neural networks. *Int. J. Remote Sens.* **1997**, *18*, 971–979. [\[CrossRef\]](#)
47. Safaei-Farouji, M.; Thanh, H.V.; Dai, Z.; Mehdodniya, A.; Rahimi, M.; Ashraf, U.; Radwan, A.E. Exploring the power of machine learning to predict carbon dioxide trapping efficiency in saline aquifers for carbon geological storage project. *J. Clean. Prod.* **2022**, *372*, 133778. [\[CrossRef\]](#)
48. Kim, M.; Gilley, J.E. Artificial Neural Network estimation of soil erosion and nutrient concentrations in runoff from land application areas. *Comput. Electron. Agric.* **2008**, *64*, 268–275. [\[CrossRef\]](#)
49. Dhaliwal, J.K.; Panday, D.; Saha, D.; Lee, J.; Jagadamma, S.; Schaeffer, S.; Mengistu, A. Predicting and interpreting cotton yield and its determinants under long-term conservation management practices using machine learning. *Comput. Electron. Agric.* **2022**, *199*, 107107. [\[CrossRef\]](#)
50. Fakoor Sharghi, A.R.; Makarian, H.; Derakhshan Shadmehri, A.; Rohani, A.; Abbasdokht, H. Predicting Spatial Distribution of Redroot Pigweed (*Amaranthus retroflexus* L.) using the RBF Neural Network Model. *J. Agric. Sci. Technol.* **2018**, *20*, 1493–1504.
51. Vakil-Baghmishah, M.-T.; Pavešić, N. Premature clustering phenomenon and new training algorithms for LVQ. *Pattern Recognit.* **2003**, *36*, 1901–1912. [\[CrossRef\]](#)
52. Azadeh, A.; Ghaderi, S.F.; Sohrabkhani, S. Forecasting electrical consumption by integration of neural network, time series and ANOVA. *Appl. Math. Comput.* **2007**, *186*, 1753–1761. [\[CrossRef\]](#)
53. Bayati, H.; Najafi, A. Performance comparison artificial neural networks with regression analysis in trees trunk volume estimation. *For. Wood Prod.* **2013**, *66*, 177–191.
54. Peng, Y.; Zhu, T.; Li, Y.; Dai, C.; Fang, S.; Gong, Y.; Wu, X.; Zhu, R.; Liu, K. Remote prediction of yield based on LAI estimation in oilseed rape under different planting methods and nitrogen fertilizer applications. *Agric. For. Meteorol.* **2019**, *271*, 116–125. [\[CrossRef\]](#)
55. Eugenio, F.C.; Grohs, M.; Venancio, L.P.; Schuh, M.; Bottega, E.L.; Ruoso, R.; Schons, C.; Mallmann, C.L.; Badin, T.L.; Fernandes, P. Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery. *Remote Sens. Appl. Soc. Environ.* **2020**, *20*, 100397. [\[CrossRef\]](#)
56. Niedbała, G.; Nowakowski, K.; Rudowicz-Nawrocka, J.; Piekutowska, M.; Weres, J.; Tomczak, R.J.; Tyksiński, T.; Pinto, A. Multicriteria prediction and simulation of winter wheat yield using extended qualitative and quantitative data based on artificial neural networks. *Appl. Sci.* **2019**, *9*, 2773. [\[CrossRef\]](#)
57. Parmley, K.A.; Higgins, R.H.; Ganapathysubramanian, B.; Sarkar, S.; Singh, A.K. Machine Learning Approach for Prescriptive Plant Breeding. *Sci. Rep.* **2019**, *9*, 17132. [\[CrossRef\]](#) [\[PubMed\]](#)
58. Abdipour, M.; Younessi-Hmazekhanlu, M.; Ramazani, S.H.R. Artificial neural networks and multiple linear regression as potential methods for modeling seed yield of safflower (*Carthamus tinctorius* L.). *Ind. Crops Prod.* **2019**, *127*, 185–194. [\[CrossRef\]](#)
59. Pandith, V.; Kour, H.; Singh, S.; Manhas, J.; Sharma, V. Performance Evaluation of Machine Learning Techniques for Mustard Crop Yield Prediction from Soil Analysis. *J. Sci. Res.* **2020**, *64*, 394–398. [\[CrossRef\]](#)
60. Abbas, F.; Afzaal, H.; Farooque, A.A.; Tang, S. Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy* **2020**, *10*, 1046. [\[CrossRef\]](#)
61. Crane-Droesch, A. Ac ce pte us Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.* **2018**, *13*, 114003. [\[CrossRef\]](#)
62. Folberth, C.; Baklanov, A.; Balkovič, J.; Skalský, R.; Khabarov, N.; Obersteiner, M. Spatio-temporal downscaling of gridded crop model yield estimates based on machine learning. *Agric. For. Meteorol.* **2019**, *264*, 1–15. [\[CrossRef\]](#)
63. Amaratunga, V.; Wickramasinghe, L.; Perera, A.; Jayasinghe, J.; Rathnayake, U.; Zhou, J.G. Artificial Neural Network to Estimate the Paddy Yield Prediction Using Climatic Data. *Math. Probl. Eng.* **2020**, *2020*, 8627824. [\[CrossRef\]](#)
64. Matsumura, K.; Gaitan, C.F.; Sugimoto, K.; Cannon, A.J.; Hsieh, W.W. Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. *J. Agric. Sci.* **2015**, *153*, 399–410. [\[CrossRef\]](#)
65. Maya Gopal, P.S.; Bhargavi, R. A novel approach for efficient crop yield prediction. *Comput. Electron. Agric.* **2019**, *165*, 104968. [\[CrossRef\]](#)
66. Rajkovi, D.; Pezo, L.; Lon, B.; Zanetti, F.; Monti, A.; Kondi, A. Yield and Quality Prediction of Winter Rapeseed—Artificial Neural Network and Random Forest Models. *Agronomy* **2022**, *12*, 58. [\[CrossRef\]](#)
67. Niedbała, G. Application of artificial neural networks for multi-criteria yield prediction of winter rapeseed. *Sustainability* **2019**, *11*, 533. [\[CrossRef\]](#)
68. Kross, A.; Znoj, E.; Callegari, D.; Kaur, G.; Sunohara, M.; Lapen, D.; 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. [\[CrossRef\]](#)
69. Alvarez, R. Predicting average regional yield and production of wheat in the Argentine Pampas by an artificial neural network approach. *Eur. J. Agron.* **2009**, *30*, 70–77. [\[CrossRef\]](#)

70. Haghverdi, A.; Washington-Allen, R.A.; Leib, B.G. Prediction of cotton lint yield from phenology of crop indices using artificial neural networks. *Comput. Electron. Agric.* **2018**, *152*, 186–197. [\[CrossRef\]](#)
71. Shahhosseini, M.; Hu, G.; Archontoulis, S.V. Forecasting corn yield with machine learning ensembles. *Front. Plant Sci.* **2020**, *11*, 1120. [\[CrossRef\]](#) [\[PubMed\]](#)
72. Hoffman, A.L.; Kemanian, A.R.; Forest, C.E. Analysis of climate signals in the crop yield record of sub-Saharan Africa. *Glob. Chang. Biol.* **2018**, *24*, 143–157. [\[CrossRef\]](#) [\[PubMed\]](#)
73. Nwosu-Obieogu, K.; Umunna, M. Rubber seed oil epoxidation: Experimental study and soft computational prediction. *Ann. Fac. Eng. Hunedoara-Int. J. Eng.* **2021**, *4*, 65–70.
74. Gulzar, Y.; Hamid, Y.; Soomro, A.B.; Alwan, A.A.; Journaux, L. A convolution neural network-based seed classification system. *Symmetry* **2020**, *12*, 2018. [\[CrossRef\]](#)
75. Albarrak, K.; Gulzar, Y.; Hamid, Y.; Mehmood, A.; Soomro, A.B. A Deep Learning-Based Model for Date Fruit Classification. *Sustainability* **2022**, *14*, 6339. [\[CrossRef\]](#)
76. Hamid, Y.; Wani, S.; Soomro, A.B.; Alwan, A.A.; Gulzar, Y. Smart Seed Classification System Based on MobileNetV2 Architecture. In Proceedings of the 2022 2nd International Conference on Computing and Information Technology (ICCIT), Tabuk, Saudi Arabia, 25–27 January 2022; pp. 217–222.
77. Yu, X.; Lu, H.; Liu, Q. Deep-learning-based regression model and hyperspectral imaging for rapid detection of nitrogen concentration in oilseed rape (*Brassica napus* L.) leaf. *Chemom. Intell. Lab. Syst.* **2018**, *172*, 188–193. [\[CrossRef\]](#)
78. Klem, K.; Křen, J.; Šimor, J.; Kováč, D.; Holub, P.; Miša, P.; Svobodová, I.; Lukas, V.; Lukeš, P.; Findurová, H.; et al. Improving nitrogen status estimation in malting barley based on hyperspectral reflectance and artificial neural networks. *Agronomy* **2021**, *11*, 2592. [\[CrossRef\]](#)
79. Berger, K.; Verrelst, J.; Féret, J.-B.; Hank, T.; Wocher, M.; Mauser, W.; Camps-Valls, G. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *92*, 102174. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Abdipour, M.; Ramazani, S.H.R.; Younessi-Hmazekhanlu, M.; Niazian, M. Modeling oil content of sesame (*Sesamum indicum* L.) using artificial neural network and multiple linear regression approaches. *J. Am. Oil Chem. Soc.* **2018**, *95*, 283–297. [\[CrossRef\]](#)
81. Zhu, H.; Chu, B.; Zhang, C.; Liu, F.; Jiang, L.; He, Y. Hyperspectral Imaging for Presymptomatic Detection of Tobacco Disease with Successive Projections Algorithm and Machine-learning Classifiers. *Sci. Rep.* **2017**, *7*, 4125. [\[CrossRef\]](#)
82. Thakur, A.; Thakur, R. Machine Learning Algorithms for Oilseed Disease Diagnosis. Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA). 2019. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3372216 (accessed on 27 August 2022).
83. Çetin, N.; Karaman, K.; Beyzi, E.; Sağlam, C.; Demirel, B. Comparative Evaluation of Some Quality Characteristics of Sunflower Oilseeds (*Helianthus annuus* L.) Through Machine Learning Classifiers. *Food Anal. Methods* **2021**, *14*, 1666–1681. [\[CrossRef\]](#)
84. Saad, P.; Ismail, N. *Artificial Neural Network Modelling of Rice Yield Prediction in Precision Farming*; Artificial Intelligence and Software Engineering Research Lab, School of Computer & Communication Engineering, Northern University College of Engineering (KUKUM): Jejawi, Perlis, 2009.
85. Mokarram, M.; Bijanzadeh, E. Prediction of biological and grain yield of barley using multiple regression and artificial neural network models. *Aust. J. Crop Sci.* **2016**, *10*, 895–903. [\[CrossRef\]](#)
86. Chen, C.; McNairn, H. A neural network integrated approach for rice crop monitoring. *Int. J. Remote Sens.* **2006**, *27*, 1367–1393. [\[CrossRef\]](#)
87. Gad, H.A.; El-Ahmady, S.H. Prediction of thymoquinone content in black seed oil using multivariate analysis: An efficient model for its quality assessment. *Ind. Crops Prod.* **2018**, *124*, 626–632. [\[CrossRef\]](#)
88. Rezazadeh Joudi, A.; Sattari, M. Estimation of scour depth of piers in hydraulic structures using gaussian process Regression. *Irrig. Drain. Struct. Eng. Res.* **2016**, *16*, 19–36.
89. Singh, A.; Ganapathysubramanian, B.; Singh, A.K.; Sarkar, S. Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends Plant Sci.* **2016**, *21*, 110–124. [\[CrossRef\]](#) [\[PubMed\]](#)
90. Bashier, I.H.; Mosa, M.; Babikir, S.F. Sesame Seed Disease Detection Using Image Classification. In Proceedings of the 2020 International Conference on Computer, Control, Electrical, and Electronics Engineering (ICCCEEE), Khartoum, Sudan, 26–28 February 2021; Volume 2021, pp. 1–5. [\[CrossRef\]](#)
91. Sahni, V.; Srivastava, S.; Khan, R. Modelling Techniques to Improve the Quality of Food Using Artificial Intelligence. *J. Food Qual.* **2021**, *2021*, 2140010. [\[CrossRef\]](#)
92. Khairunniza Bejo, S.; Mustaffha, S.; Khairunniza-Bejo, S.; Ishak, W.; Ismail, W. Application of Artificial Neural Network in Predicting Crop Yield: A Review Spectroscopy techniques View project Application of Artificial Neural Network in Predicting Crop Yield: A Review. *J. Food Sci. Eng.* **2014**, *4*, 1–9.
93. Sarvestani, N.S.; Rohani, A.; Farzad, A.; Aghkhani, M.H. Modeling of specific fuel consumption and emission parameters of compression ignition engine using nanofluid combustion experimental data. *Fuel Process. Technol.* **2016**, *154*, 37–43. [\[CrossRef\]](#)
94. Sarbu, C.; Pop, H. Principal component analysis versus fuzzy principal component analysis. *Talanta-Oxf. Amst.* **2005**, *65*, 1215–1220. [\[CrossRef\]](#)
95. Ilin, A.; Raiko, T. Practical approaches to principal component analysis in the presence of missing values. *J. Mach. Learn. Res.* **2010**, *11*, 1957–2000.

96. Hoang, N.-D.; Pham, A.-D.; Nguyen, Q.-L.; Pham, Q.-N. Estimating compressive strength of high performance concrete with Gaussian process regression model. *Adv. Civ. Eng.* **2016**, *2016*, 2861380. [[CrossRef](#)]
97. Yang, Y.-K.; Sun, T.-Y.; Huo, C.-L.; Yu, Y.-H.; Liu, C.-C.; Tsai, C.-H. A novel self-constructing Radial Basis Function Neural-Fuzzy System. *Appl. Soft. Comput.* **2013**, *13*, 2390–2404. [[CrossRef](#)]
98. Tatar, A.; Shokrollahi, A.; Mesbah, M.; Rashid, S.; Arabloo, M.; Bahadori, A. Implementing Radial Basis Function Networks for modeling CO₂-reservoir oil minimum miscibility pressure. *J. Nat. Gas Sci. Eng.* **2013**, *15*, 82–92. [[CrossRef](#)]
99. Ashtiani, S.-H.M.; Rohani, A.; Aghkhani, M.H. Soft computing-based method for estimation of almond kernel mass from its shell features. *Sci. Hortic.* **2020**, *262*, 109071. [[CrossRef](#)]
100. Zareei, J.; Rohani, A. Optimization and study of performance parameters in an engine fueled with hydrogen. *Int. J. Hydrogen Energy* **2020**, *45*, 322–336. [[CrossRef](#)]
101. Jiang, P.; Chen, J. Displacement prediction of landslide based on generalized regression neural networks with K-fold cross-validation. *Neurocomputing* **2016**, *198*, 40–47. [[CrossRef](#)]
102. Rahimi, M.; Pourramezan, M.-R.; Rohani, A. Modeling and classifying the in-operando effects of wear and metal contaminations of lubricating oil on diesel engine: A machine learning approach. *Expert. Syst. Appl.* **2022**, *203*, 117494. [[CrossRef](#)]
103. Shim, K.B.; Shin, S.H.; Shon, J.Y.; Kang, S.G.; Yang, W.H.; Heu, S.G. Classification of a collection of sesame germplasm using multivariate analysis. *J. Crop Sci. Biotechnol.* **2016**, *19*, 151–155. [[CrossRef](#)]
104. Fiseha, B.; Yemane, T.; Fetien, A. Assessing inter-relationship of sesame genotypes and their traits using cluster analysis and principal component analysis methods. *Int. J. Plant Breed Genet.* **2015**, *9*, 228–237.
105. Ismaila, A.; Usman, A. Genetic Variability for Yield and Yield Components in Sesame (*Sesamum indicum* L.). *Electron. J. Plant Breed.* **2014**, *3*, 2012–2015.
106. Yol, E.; Karaman, E.; Furat, S.; Uzun, B. Assessment of selection criteria in sesame by using correlation coefficients, path and factor analyses. *Aust. J. Crop Sci.* **2010**, *4*, 598–602.
107. Sudret, B. Global sensitivity analysis using polynomial chaos expansions. *Reliab. Eng. Syst. Saf.* **2008**, *93*, 964–979. [[CrossRef](#)]
108. Saltelli, A.; Ratto, M.; Andres, T.; Campolongo, F.; Cariboni, J.; Gatelli, D.; Saisana, M.; Tarantola, S. *Global Sensitivity Analysis: The Primer*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
109. Wan, H.-P.; Ren, W.-X.; Todd, M.D. Arbitrary polynomial chaos expansion method for uncertainty quantification and global sensitivity analysis in structural dynamics. *Mech. Syst. Signal. Process.* **2020**, *142*, 106732. [[CrossRef](#)]