


Review

# Machine Learning for Plant Breeding and Biotechnology

Mohsen Niazian <sup>1,\*</sup>  and Gniewko Niedbała <sup>2,\*</sup> 

<sup>1</sup> Field and Horticultural Crops Research Department, Kurdistan Agricultural and Natural Resources Research and Education Center, Agricultural Research, Education and Extension Organization (AREEO), Jam-e Jam cross way, P.O. Box 741, 66169-36311 Sanandaj, Iran

<sup>2</sup> Department of Biosystems Engineering, Faculty of Environmental Engineering and Mechanical Engineering, Poznań University of Life Sciences, Wojska Polskiego 50, 60-627 Poznań, Poland

\* Correspondence: mniazian@ut.ac.ir (M.N.); gniewko@up.poznan.pl (G.N.)

Received: 18 August 2020; Accepted: 25 September 2020; Published: 27 September 2020



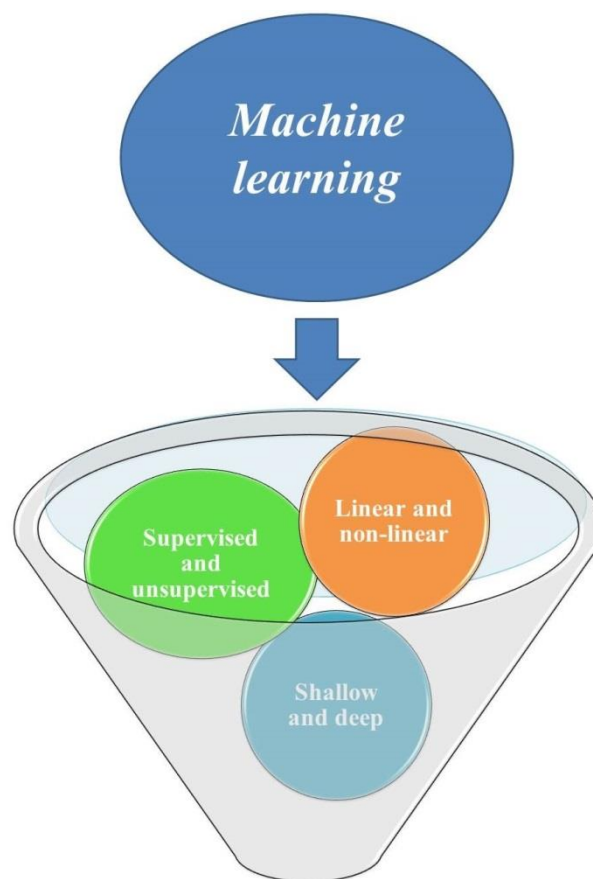
**Abstract:** Classical univariate and multivariate statistics are the most common methods used for data analysis in plant breeding and biotechnology studies. Evaluation of genetic diversity, classification of plant genotypes, analysis of yield components, yield stability analysis, assessment of biotic and abiotic stresses, prediction of parental combinations in hybrid breeding programs, and analysis of in vitro-based biotechnological experiments are mainly performed by classical statistical methods. Despite successful applications, these classical statistical methods have low efficiency in analyzing data obtained from plant studies, as the genotype, environment, and their interaction ( $G \times E$ ) result in nondeterministic and nonlinear nature of plant characteristics. Large-scale data flow, including phenomics, metabolomics, genomics, and big data, must be analyzed for efficient interpretation of results affected by  $G \times E$ . Nonlinear nonparametric machine learning techniques are more efficient than classical statistical models in handling large amounts of complex and nondeterministic information with “multiple-independent variables versus multiple-dependent variables” nature. Neural networks, partial least square regression, random forest, and support vector machines are some of the most fascinating machine learning models that have been widely applied to analyze nonlinear and complex data in both classical plant breeding and in vitro-based biotechnological studies. High interpretive power of machine learning algorithms has made them popular in the analysis of plant complex multifactorial characteristics. The classification of different plant genotypes with morphological and molecular markers, modeling and predicting important quantitative characteristics of plants, the interpretation of complex and nonlinear relationships of plant characteristics, and predicting and optimizing of in vitro breeding methods are the examples of applications of machine learning in conventional plant breeding and in vitro-based biotechnological studies. Precision agriculture is possible through accurate measurement of plant characteristics using imaging techniques and then efficient analysis of reliable extracted data using machine learning algorithms. Perfect interpretation of high-throughput phenotyping data is applicable through coupled machine learning-image processing. Some applied and potentially applicable capabilities of machine learning techniques in conventional and in vitro-based plant breeding studies have been discussed in this overview. Discussions are of great value for future studies and could inspire researchers to apply machine learning in new layers of plant breeding.

**Keywords:** artificial neural networks; big data; classification; high-throughput phenotyping; modeling; predicting

## 1. Introduction

Due to climate change (global warming), increasing food requirements and depletion of resources in consequence of increasing global population, it is necessary to use modern technologies in agriculture and food sciences [1]. Plant breeding is a dynamic branch of agricultural science. It started with simple selection of impressive plants with superior characteristics. Later, genetics and statistics were involved in classical plant breeding, mainly after the discoveries of Gregor Mendel and Sir Ronald Aylmer Fisher. Next, modern plant breeding emerged with the advancements in genetic and biotechnology approaches. Classical plant breeding methods mainly included assessment and classification of genetic diversity, yield components analysis (indirect selection of superior genotypes with impressive economic characteristics), yield stability analysis (genotype  $\times$  environment interaction), enhanced tolerance to biotic and abiotic stresses, and hybrid breeding programs. In vitro-based biotechnological breeding methods mainly included in vitro micropropagation, doubled haploid production, artificial polyploidy induction, and *Agrobacterium*-mediated gene transformation. In in vitro micropropagation studies, researchers want to investigate the effects of influential factors (inputs), such as combination of culture medium components, combination and concentrations of plant growth regulators (PGRs), and interactions of plant genotype  $\times$  culture medium  $\times$  PGRs  $\times$  explant type  $\times$  explant age  $\times$  elicitor additives  $\times$  type and concentration of carbohydrate source  $\times$  etc., on regeneration efficiency (outputs) of their desired plants. Classical statistical techniques have been employed to analyze and interpret the results of both classical and in vitro-based plant breeding studies. These analytical techniques are mainly based on variance and linear regression models to assess the relationship of variables and predict the effect of independent variables on dependent variables. One regression model is required to assess the effect of a group of independent variables ( $X_1, X_2, X_3, \dots, X_n$ ) on one dependent variable ( $Y$ ), according to the multiple linear relationships [2]. However, nonlinear and nondeterministic properties are inextricably linked with plant biological systems [3]. Therefore, despite of successful applications, the classical linear regression-based models are unable to interpret highly nonlinear and complex relationships between dependent and independent variables. Most of these plant breeding approaches are “multiple-independent variables versus multiple-dependent variables.” Under these conditions, one regression model is required for each output [4]. Powerful data mining tools are employed in plant breeding studies to predict and explain complex data.

Machine learning—the science of programming computers so they can learn from data—has been widely applied in both classical and in vitro-based plant breeding studies to interpret the flow of information about plants from the DNA sequence to the observed phenotypes. There are three ways to classify machine learning methods, including supervised and supervised models, linear and nonlinear algorithms, and shallow and deep learning models (Figure 1). Artificial neural networks (ANNs), deep neural networks (DNNs), convolutional neural networks (CNNs), random forest (RF), and support vector machines (SVMs) are examples of nonlinear nonparametric machine learning algorithms, applied for processing nonlinear data in plant studies [5]. These data-driven models are able to parse and interpret non-normal, nonlinear, and nondeterministic unpredictable data sets, through the full use of all spectral data and avoid irrelevant spectral bands and multicollinearity [6,7]. Among different learning algorithms, including supervised, unsupervised, reinforcement, sparse dictionary, and rule-based, supervised learning is more suitable and efficient for life science problems [8]. Supervised learning can be used for classification (predicting non-numeric answers) and regression (predicting numeric answers) [9]. Formless datasets such as data obtained by photo imaging or sequencing can be interpreted through machine learning algorithms [10]. Genome sequencing data can be used in machine learning models for the identification and classification of transposable elements [11]. By using machine learning algorithms, breeders are able to predict multiple outputs (multiple-dependent variables) through different combinations of multiple inputs in one model and reduce required analyses.



**Figure 1.** Different categories of machine learning algorithms.

Artificial neural networks, consist of an input, an output, and several hidden layers, are nonlinear nonparametric models which do not require a prior structure for data and detailed information about the physical processes to be modeled and to tolerate data loss [12,13]. Because of their more hidden layers, DNNs have greater predictive power than ANNs. Convolutional neural networks, as state-of-the-art deep learning architecture, are inspired by the natural visual perception mechanism of the living creatures and consist of convolutional, pooling, fully-connected layers, and an output layer [14]. CNNs are suitable for classification studies because of automatic feature extraction [9]. Image classification, object detection, object tracking, pose estimation, text detection and recognition, visual saliency detection, action recognition, scene labeling, speech, and natural language processing are some of the typical applications of CNNs [14]. Neural networks have low interpretability of the features (lack the interpretation capability), especially CNN in which the features extracted are hidden. More advanced machine learning technique of SVMs, which uses a supervised learning algorithm to find both linear and nonlinear relationships in data, can be used for clustering, classification, and regression analysis of data sets. In comparison with multilayer perceptron (MLP) of ANN, SVM uses a large number of hidden units and has better performance in the formulation of the learning problem, subsequently quadratic optimization task [15]. Random forest regression is a regression tree-based machine learning that uses multiple decision trees to classify data and needs setting the number of trees, the number of random features, and the stop criteria for training. RF is more suitable for spectral data analysis and overfitting can be controlled through combining different independent predictors [16,17]. In semantic segmentation methods, such as automated phenotyping and plant disease detection, deep learning CNN can be more effective than shallow learning models of SVMs and RF and problem of required large manually crafted features can be solved by using image augmentation and small manually annotated empirical dataset for fine-tuning a synthetically

bootstrapped CNN [18]. Through the integrating image feature extraction with classification in a single pipeline, deep convolutional neural networks have been considered as mainstream in biotic and abiotic stress diagnosis and classification [19]. A nine-layer deep CNN model was trained for identification of plant leaf diseases using data set with 39 different classes of plant leaves and background images and 96.46% classification accuracy was reported, which is greater than traditional machine learning approaches of SVM, decision tree, logistic regression, and K-NN [20]. CNNs are also applicable in remote sensing for object detection and pattern recognition. High accuracy (84%) for fine-grained mapping of vegetation species and communities using deep CNN-based segmentation, trained by data directly derived from visual interpretation of unmanned aerial vehicles (UAV)-based high-resolution Red-Green-Blue (RGB) imagery, has been reported [21].

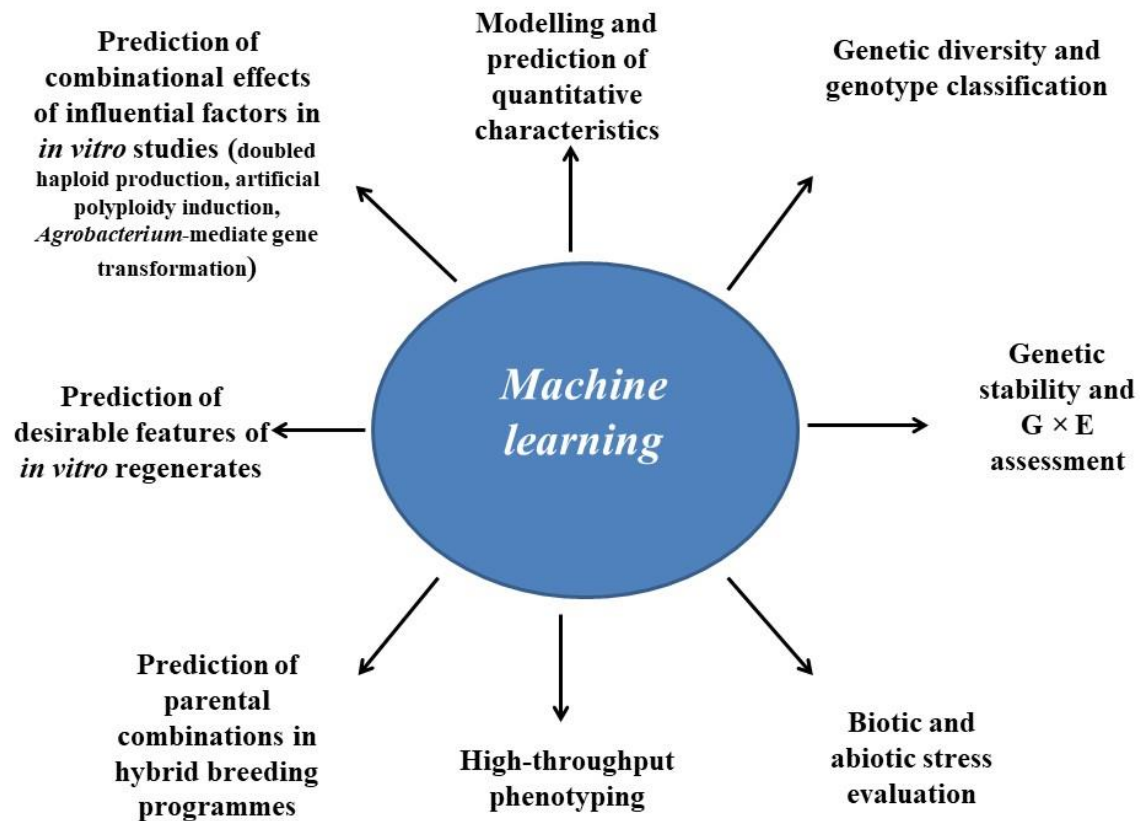
A lot of training data is required in ANN for the optimization of sigmoid functions belonging to the hidden layer's neurons, as overfitting and local minima may happen by small number of training data. Therefore, the optimization process cannot be properly carried using back-propagation algorithms, when the number of training samples is small [8]. Through the short review on studies that used SVM and ANN techniques for identifying disease in plants, it was concluded that the ANN-based methods are better than SVM-based methods, as few samples and features are used in SVM-based methods to identify the disease-affected plants [22]. Conversely, in modeling in vitro culture of *Chrysanthemum* (*Dendranthema × grandiflorum*), better performance accuracy of SVR ( $R^2 > 0.92$ ) than MLP ( $R^2 > 0.82$ ) has been reported [15]. Applying different algorithm and comparing their performance is an appropriate solution to find the best algorithm in a particular data set. In tea plant (*Camellia sinensis* L.), partial least squares discriminative analysis (PLS-DA) and least squares-support vector machines (LS-SVM) were used for the classification of different nitrogen nutrition status under field condition and better performance with correct classification of LS-SVM than PLS-DA was reported [23]. The pros and cons of different nonlinear machine learning methods under similar scenarios are presented in Table 1.

**Table 1.** Pros and cons of nonlinear machine learning algorithms applied in classical and in vitro-based plant breeding studies.

Learning Algorithm	Advantages	Disadvantages
ANNs	<ul style="list-style-type: none"> <li>• Good learning capabilities</li> </ul>	<ul style="list-style-type: none"> <li>• Lack the interpretation capability</li> <li>• Overfitting and local minima in small number of training data</li> <li>• Implementing very small number of hidden neurons</li> </ul>
CNNs	<ul style="list-style-type: none"> <li>• Ability of automatic feature extraction</li> </ul>	<ul style="list-style-type: none"> <li>• Lack the interpretation capability</li> <li>• Require large amounts of data for training</li> <li>• Require considerable skill and experience to select suitable hyperparameters</li> </ul>
SVMs	<ul style="list-style-type: none"> <li>• Uses a large number of hidden units</li> <li>• Quadratic optimization task in the formulation of the learning problem</li> </ul>	<ul style="list-style-type: none"> <li>• Shallow architecture</li> </ul>
RF	<ul style="list-style-type: none"> <li>• Ability to handle noise</li> <li>• Prevent overfitting</li> <li>• Ability to manage a large number of features</li> </ul>	<ul style="list-style-type: none"> <li>• Shallow architecture</li> </ul>

ANNs—artificial neural networks; CNN—convolutional neural networks; RF—random forest; SVMs—support vector machines.

Different application areas for nonlinear machine learning technologies in classical and in vitro-based plant breeding studies are shown in Figure 2. The following sections of the article provide a comprehensive review of the applications of these nonlinear machine learning techniques in classical and in vitro-based plant breeding studies.



**Figure 2.** Potential applications of machine learning techniques in classical and modern plant breeding.

Some recently applied nonlinear machine learning models in both classical and in vitro-based plant breeding studies are listed in Table 2.

**Table 2.** Examples of recently applied nonlinear machine learning models in classical and modern plant breeding studies.

Plant Species	Type of Machine Learning	Techniques	Purpose(s)	Reference
Ajowan ( <i>Trachyspermum ammi</i> L.)	ANN	MLR	Modeling and predicting of seed yield	[2]
	ANN	MLR	Modeling and predicting of essential oil content	[24]
	ANN	MLR, IP	Predicting physical properties of embryogenic callus and number of somatic embryos	[25]
<i>Arabidopsis thaliana</i>	DT, SVMs, NB	Gaussian kernel	Predict the plant abiotic stresses response through the miRNAs' concentration	[8]
Carrot ( <i>Daucus carota</i> )	RF	-	Precision agriculture-yield mapping	[26]
Chrysanthemum	ANN	GA	Modeling and optimizing of in vitro sterilization	[27]
	ANFIS	GA	Modeling and optimizing of somatic embryogenesis	[3]
Cucumber ( <i>Cucumis sativus</i> )	ANN, SVMs	MLP	Modeling effect of plant growth regulators on somatic embryogenesis	[15]
	CNN	IP	Segmentation and quantification of powdery mildew disease	[28]
Garnem (G × N15) <i>Prunus</i> rootstock	ANN	GA	Prediction and optimization of mineral salts of in vitro culture medium	[29]
	ANN	GA	Modeling and optimizing of in vitro hormonal combination	[30]
	ANN	GA	Modeling and optimizing of new in vitro culture medium	[31]
Grapevine rootstock	ANN	Principal coordinate analysis, UPGMA	Genetic diversity assessment through molecular markers (RAPD-SSR) dataset	[32]
Maize ( <i>Zea mays</i> L.)	CNN	IP	Identification of haploid and diploid maize seeds	[33]
	CNN	IP	Classification model to identify the infected and healthy leaves	[34]
	CNN	IP	Plant diseases recognition	[35]
	CNN	IP	Identification and classification of drought stress	[19]
Okra ( <i>Abelmoschus esculentus</i> L.)	DNN	IP	High-throughput salt-stress phenotyping	[36]
Pearl millet ( <i>Pennisetum glaucum</i> )	DNN	IP	Identification of mildew disease	[37]
Potato ( <i>Solanum tuberosum</i> )	ANN	IP	Identification and discrimination of potato varieties	[38]
	RF		Classification of <i>Phytophthora infestans</i> infected cultivars	[17]
Rapeseed ( <i>Brassica napus</i> )	ANN	MLP	Seed yield modeling	[39]
	CNN	IP	Stand count estimation	[40]
	ANN	MLP	Multicriteria yield prediction based on meteorological data and mineral fertilization data	[41]
	ANN	MLP	Early prediction and simulation of seed yield based on meteorological and mineral fertilization data	[42]
Rice ( <i>Oryza sativa</i> )	CNN		Plant diseases and pest recognition	[43,44]
Safflower ( <i>Carthamus tinctorius</i> L.)	ANN	MLR	Seed yield modeling	[45]
Sesame ( <i>Sesamum indicum</i> L.)	ANN	MLR	Oil content modeling	[46]
	ANN, SVMs	RBF, ERBF, GRNN, M5-Rule, M5-Tree, MLR	Estimation of oil and protein content	[47]
Soybean ( <i>Glycine max</i> )	CNN	IP	Estimation of seeds per pod	[48]
	DNN	IP	Evaluation of stomatal density diversity	[49]
Tomato ( <i>Lycopersicon esculentum</i> L.)	ANN	MLR, IP	Modeling of callus induction and regeneration in anther culture	[50]
	CNN	IP	Evaluation of disease severity	[51]
	ANN	MLP	Estimation of salinity tolerance	[52]
	ANN	MLP	Prediction of seed yield based on meteorological data and information on mineral fertilization	[53]
	ANN	MLP	Prediction and simulation of seed yield with qualitative and quantitative data sets	[54]
Wheat ( <i>Triticum aestivum</i> L.)	CNN	IP	Quantification of spikes	[55]
	DNN	LSTM	Production forecasting	[56]
	CNN	-	Genomic selection	[57]
	ANN, GRNN	MLP	Modeling in vitro shoot regeneration	[58]
White ginger ( <i>Hedychium coronarium</i> )	ANN	MLP	Analysis of concentration of ferulic acid, deoxynivalenol, and nivalenol	[59]
	ANN	MLP	Prediction and optimization of coronarin D content	[60]

ANFIS—adaptive neuro-fuzzy inference system; ANN—artificial neural networks; CNN—convolutional neural networks; DNN—deep neural network; DT—decision tree; ERBF—extended radial bases function; GA—genetic algorithm; GRNN—generalized regression neural network; IP—image processing; LSTM—long short term memory; MLR—multiple linear regressions; MLP—multilayer perceptron; NB—Naïve Bayes; RBF—radial bases function; RF—random forest; SVMs—support vector machines; UPGMA—unweighted pair group method with arithmetic mean.



## 2. Traditional Plant Breeding

### 2.1. Assessment and Classification of Genetic Diversity

One of the most important prerequisites of plant breeding programs is genetic diversity, which enables selection of important accessions and their use in future breeding programs [61]. Morphological, biochemical, and physiological markers have been analyzed to investigate the genetic diversity of different plants. Morphological features are the simplest to measure and do not require special tools or techniques. The statistical analysis of these markers can prove the existence of genetic diversity among studied genotypes. Niazi et al. [62] used analysis of variance (ANOVA) and estimated the coefficient of variation (CV) of different agro-morphological traits (plant height, number of branches, number of umbels, number of umbellets in an inflorescence, biological yield, and single plant yield) of eight ecotypes of ajowan medicinal plant (*Carum copticum* L.) and observed significant genetic diversity.

Molecular/genetic markers are another group of markers which enable assessment of genetic diversity and discrimination of the genotype. Amplified fragment-length polymorphism (AFLP), restriction-fragment length (RFLP), randomly amplified polymorphic DNA (RAPD), simple sequence repeat (SSR), intersimple sequence repeats (ISSR), and single nucleotide polymorphism (SNP) are the most commonly used molecular markers to study genetic diversity and species identification in different target plants [63]. These genetic markers estimate phylogenetic relationships and identify varieties more reliably and effectively than morphological markers [64]. Although molecular markers are more effective than morphological markers in the assessment of genetic diversity and discrimination and identification of various plant genotypes, there are some technical and/or economic limitations [65].

Classical multivariate analyses such as cluster analysis, discriminant function analysis, and principal component analysis (PCA) have been used for the classification and grouping of different genotypes in various plant species by means of morphological, biochemical, physiological, and molecular markers [61,64,66–68]. Object detection through deep learning algorithms could be used for efficient genetic diversity assessment and classification of plant genotypes. The use of CNN to classify morphological parameters is an appropriate alternative to conventional classification methods, such as k-nearest neighbor, probabilistic neural network, support vector machine, genetic algorithm, and PCA, all of which are time consuming and require feature extraction [65,69]. In soybeans (*Glycine max* (L.) Merr.), the genetic diversity of 90 accessions was detected through high-throughput evaluation of stomatal density [49]. In *Cinnamomum osmophloeum* Kanehira (Lauraceae), deep CNN was applied for differentiating between morphologically similar species, and accuracy of CNN classifiers was better than SVMs classifiers (96.7% vs. 74.6%) [70]. Sant'Anna et al. [71] compared the performance of ANN with Fisher's classical multivariate statistical technique and Anderson's discriminant functions to assess the genetic diversity and classify 10 plant populations. They observed that ANN-classified populations with high and low differentiation were better than classical methods, as there were fewer wrongly classified individuals. Linear discriminant analysis and nonlinear artificial neural network methods were applied to identify and discriminate 10 potato varieties with morphological data obtained through image processing. The correctness of classification of the ANN method was 100% [38].

As was mentioned above, machine learning can also be used for classification through molecular markers data. DNA/RNA sequences can be used for training CNNs and applications in plant molecular biology and classification of genotypes through molecular markers [9]. Different machine learning models were used to identify true single nucleotide polymorphisms (SNPs) in allopolyploid peanuts (*Arachis hypogaea* L.). The selection of true SNPs by means of real peanut RNA sequencing (RNA-seq) and whole-genome shotgun (WGS) resequencing data resulted in above 80% accuracy [72]. Costa et al. [32] applied a neural network algorithm to infer the genetic diversity and group allelic frequencies obtained by RAPD and SSR molecular markers in grapevine rootstock varieties and

found three genetically diversified clusters among 64 grapevine rootstocks analyzed. Deep learning techniques enable prediction of plant phenotypes from their genome data [9].

Artificial neural networks have also been applied for genomic prediction and genomic selection in different plant species [73–75]. The phenotypes of 2000 Iranian bread wheat landrace accessions were predicted from genomic dataset collected from 33,709 DArT markers using a deep convolutional neural network. Authors reported that the Pearson's correlation coefficients between observed and predicted phenotypic values (grain length, grain width grain hardness, thousand-kernel weight, test weight, sodium dodecyl sulfate sedimentation, grain protein, and plant height) in deep CNN were more than other genomic selection methods [57].

## 2.2. Yield Component Analysis and Indirect Selection (Prediction)

An increase in the economic yield (seed yield, oil yield, sugar yield, essential oil yield, biomass yield, straw yield, lint percentage, etc.) is always the target of most breeding programs. However, yield is a highly complex quantitative trait, which is usually controlled by several genes, and it is strongly influenced by the environment. Therefore, yield traits have low heritability and direct selection does not improve such complex traits. Instead, plant breeders usually prefer to use simpler highly correlated traits to have greater influence on yield. Selected yield component(s) will be used as “selection criteria” in future studies, i.e., indirect selection [2,76]. Classical single variable and multivariate linear methods, such as correlation coefficient analysis, PCA, path analysis, and multiple regression analyses (stepwise, forward, and reverse), have been used in classical plant breeding to interpret relationships between plant traits and improve important quantitative properties like yield and tolerance to biotic and abiotic stresses. The correlation coefficient analysis and path analysis have been used to evaluate a simple relationship between two traits and identify cause/effect relationships between correlated variables, respectively [24]. Regression-based methods are the most effective multivariate statistical methods for indirect selection purposes. They are based on a linear relationship of a dependent variable (Y) as a function of multiple independent variables. These multiple variables create a complicated condition for interpretation. However, some reduction techniques, such as PCA and factor analysis, are able to concentrate the original multiple variables in a few complex variables [77].

Stepwise, forward, and reverse regression analyses have been used to determine the effects of yield components on different economic quantitative characteristics in various crops. Backward stepwise regression was used to find the relationship between changes in grain yield and yield components of rice (*Oryza sativa* L.) in terms of the relative response ratio to elevated CO<sub>2</sub> [78]. Stepwise regression was used to determine the components of sugar beet (*Beta vulgaris* L.) yield affecting the yield of sugar under water deficit regimes and foliar application of jasmonic acid [76]. Zou et al. [77] applied stepwise regression analysis to identify the yield components involved in drought resistance of cotton seedlings (*Gossypium hirsutum* L.). Despite all the advantages, there is one major drawback to regression-based models in classical plant breeding studies—it is impossible to analyze nonlinear relationships of dependent and independent variables [2,79]. The application of nonlinear machine learning algorithms in yield component analysis and indirect selection studies enables the interpretation of nonlinear relationships between dependent and independent variables, the contribution of yield components to yield and prediction of economic quantitative characteristics. ANN was more efficient than multiple linear regressions (MLR) in the prediction of seed yield [2] and essential oil content [24] of ajowan (*Trachyspermum ammi* L.). Emamgholizadeh et al. [79] found that ANN predicted the yield of sesame seeds (*Sesamum indicum* L.) better than MLR. The ANN model was characterized by lower root mean square error (RMSE) and higher determination coefficient (R<sup>2</sup>). The analysis of the sensitivity of the ANN model showed that the number of capsules per plant and the flowering time were, respectively, the most and the least significant variables to the yield of sesame seeds. Artificial neural networks have successfully predicted the yield of apples, pears, chives, and onions, allowing for data on crop diseases, time until harvest (based on the date), current temperature, humidity and precipitation (amount of snowfall) in the area, amount of sunshine, ground temperature, atmospheric pressure, and moisture



evaporation in the ground [80]. ANNs have also been used to predict the yield of winter rapeseed and winter wheat on the basis of meteorological data (air temperature and precipitation) and information on the use of mineral fertilizer [41,42,53,54]. The superiority of DNN (Long Short-Term Memory) over the auto regressive integrated moving average (ARIMA) model in predicting wheat production has been reported [56]. Deep CNN classification has been applied for image-based ear counting of wheat with high level of robustness, without considering of variables, such as growth stage and weather conditions [55].

Neural networks have also been used to estimate and predict the qualitative characteristics of different plants. The ANN model predicted the oil content in sesame more accurately and efficiently than the MLR model [46]. Parsaeian et al. [47] applied a multilayer perceptron (MLP)-ANN to estimate the oil and protein content in sesame on the basis of 138 morphological features measured in 125 sesame seed genotypes. The morphological characteristics of seeds were measured through image processing. The qualitative parameters of oil and the protein content in sesame seeds estimated by means of  $R^2$  and RMSE statistics revealed the superiority of MLP over the radial basis function (RBF), extended RBF (ERBF), GRNN, M5-Rule, M5-Tree, support vector machine regression, and linear regression models [47]. Niedbala et al. [59] developed a multilayer perceptron ANN to assess the influence of the cultivar and weather conditions on the concentration of ferulic acid and correlate its content with the concentration of deoxynivalenol and nivalenol in 23 winter wheat genotypes with different Fusarium resistance. Independent variables consisted of 14 features, including 12 quantitative data and 2 qualitative data. The sensitivity analysis of neural networks showed that the experiment variant and winter wheat cultivar were the most important determinants of the concentration of ferulic acid, deoxynivalenol, and nivalenol in winter wheat seeds [59]. Ray et al. [60] applied an MLP-ANN model to assess the effects of topographic, soil, and environmental factors (18 input parameters, including soil nutrients and climate factors) on the content of active constituent of coronarin D in white ginger lily (*Hedychium coronarium*). The sensitivity analysis of the ANN showed that altitude, manganese, and zinc were the most important variables predicting the coronarin D content.

### 2.3. Yield Stability and Genotype $\times$ Environment Interaction

The environment (climate and soil), agricultural operations (sowing-cultivation-harvesting), and plant genotype are the factors that affect the yield and productivity of crops. The relationships (direct and/or indirect) and interactions between these factors create a complex situation determining the potential yield of plants [39]. Environmental variations and the genotype  $\times$  environment interaction (GEI) are the factors that cause year to year variations in the yield and phenotypic trait of a specific genotype. The choice of a genotype for a target trait is a complex and difficult task because of the GEI, as different genotypes respond differently to varied environmental conditions. The estimation of relative performance of genotypes over the environments, through stability analysis is a perfect solution to these yearly variations [81]. Finlay and Wilkenson's regression analysis and coefficient [82]; Eberhart and Russel's coefficient of regression ( $S^2_{di}$ ) [83]; Wricke's ecovalence ( $Wi$ ) [84], Shukla's procedure of stability variance [85], coefficient of variance (CV) [86], and Lin and Binns cultivar performance measure [87] are classical univariate approaches used for the assessment of the GEI. Linear regression analysis and variance components are the main aspects of these methods [88]. Apart from the aforementioned statistic methods, the sustainable yield index (SYI) [89] is used to evaluate the effects of agricultural practices on crop yield sustainability [90]. All these methods are parametric, and therefore, the assumptions of the distribution of data and the homogeneity of variance should be considered before they are applied [91]. There are nonparametric univariate methods to evaluate the GEI, including  $S_i^1$ ,  $S_i^2$ ,  $S_i^3$ , and  $S_i^6$  stability parameters [92,93], Kang parameter [94], Ketrunk and Ketyield plots [95], Fox-rank [96], and Star [91]. These nonparametric stability statistics are analytical clustering procedures that determine the stability of genotypes on the basis of ranks rather than data and free from modeling assumptions. A genotype is considered stable if its ranking is relatively constant across environments [91].

Principal component analysis, cluster analysis, additive main effects and multiplicative interactions (AMMI), and genotype plus genotype  $\times$  environment interaction biplot (GGE) are multivariate procedures enabling examination of multidirectional aspects of the GEI by imaging the response of a genotype in an  $E$ -dimensional space [91]. Multivariate stability analyses are more powerful and precise than univariate approaches. However, these are complex methods that do not provide a simple measure of yield stability for a reliable ranking of genotypes. Limited access to software is another bottleneck of these methods [91]. In a recent study, both linear and nonlinear regression models were applied to estimate the influence of climate variables (precipitation, sunshine duration, average relative humidity, maximum temperature, minimum temperature, and average temperature) on the growth and yield-related characteristics of cotton (the cotton height at the flowering stage, stalk weight, yield of cotton seeds, and lint percentage). The authors found that the interpretation of linear regression equations was generally lower than the interpretation of nonlinear equations [97]. There was a linear relationship or a relatively complex nonlinear relationship between the cotton growth indicators and climate variables in one site of their study, but they did not find the best equations for the cotton growth indices and the influence of climate variables on the cotton growth indices at several sites. In addition, the authors developed one regression model for each condition [97]. When several independent variables and several dependent variables are of interest, i.e., multiple-independent variables versus multiple-dependent variables, ANN can reduce the required analyses and result in higher accuracy [25]. It is clear that an ANN model can find the best equations in all studied environments in a faster and more precise manner by considering other factors such as soil and cotton properties. Plant growth indices and climate variables could be entered into an ANN model as dependent and independent variables, respectively. Then, linear and nonlinear relationships between the variables can be considered through powerful ANN models. There are well-recognized statistical and biological limitations to regression approaches. ANN modeling would enable breeders to evaluate the GEI and genetic stability of a large number of genotypes faster and more precisely. Coupled artificial intelligence (ANN) with deep phenotyping is a valuable tool for understanding plant–environment interactions [98].

#### 2.4. Biotic and Abiotic Stress Assessment

Plants are exposed to various biotic and abiotic stresses. Different approaches have been applied to assess the tolerance and resistance of plant genotypes to these stresses and to identify superior genotypes.

There have been numerous breeding attempts to combat drought stress. Plants' tolerance to drought has been studied through some statistical indices, such as tolerance (TOL), mean productivity (MP), stress susceptibility index (SSI), geometric mean productivity (GMP), harmonic mean (HARM), relative drought index (RDI), stress tolerance index (STI), yield index (YI), yield stability index (YSI), and modified stress tolerance index (K1STI and K2STI) [99–106]. These classical approaches are based on morphological data, mainly yield generated under nonstress ( $Y_p$ ) and stress ( $Y_s$ ) conditions. However, apart from morphological attributes, there are many physiological and biochemical pathways involved in plants' response to environmental stresses. Secondary metabolites, cellular antioxidants, plant growth regulators, compatible solutes, and polyamines are all involved in plants' response to biotic and abiotic stresses [107,108]. Combining phenomic data with metabolomic and genomic data is an efficient strategy to assess plants' responses to biotic and abiotic stresses [109]. Classical multivariate statistical methods are not efficient enough to manage such a large volume of data (multiple independent variables versus multiple dependent variables). Linear regression is the most common technique used for the detection of nutrition deficiencies through the RGB image technique. However, features extracted from digital images with nonlinear relationship with nutrient content cannot be explained through linear regression model [110]. Machine learning techniques, along with digital images, could be used to model and predict genotypes' responses to stressful conditions and find the ones that are more resistant to stress and nonstress environments by analyzing all phenomic and omics (metabolomic and genomic) data. Big data—imaging and remote-sensing data—can be

interpreted through machine learning for high-throughput stress phenotyping [111]. Ravari et al. [52] applied an MLP-ANN and the TOL, MP, GMP, HM, SSI, STI, YI, and YSI indices to predict the salinity tolerance of 41 Iranian wheat cultivars (*Triticum aestivum* L.). They found that the YSI, MP, GMP, and STI were the best predictors of salinity-tolerant cultivars. In *Arabidopsis* (*Arabidopsis thaliana*), miRNA expressions were used as input features to predict plant responses to abiotic stresses of drought, salinity, cold, and heat using machine learning models of decision tree (DT), SVM, least-square support vector machines (LSSVM), and Naïve Bayes (NB). It was concluded that miRNA-169, miRNA-159, miRNA-396, and miRNA-393 had the highest contributions to plant response towards abiotic stresses and the SVM with Gaussian kernel had better performance than other machine learning methods in prediction of plant stress response ( $R^2 = 0.96$ ) [8]. Deep CNN along with traditional machine learning method was used for identification and classification of maize drought stress through the field-obtained data under optimum moisture, light drought, and moderate drought stress. Authors reported identification accuracy of 98.14%, which was more than Gradient Boosting Decision Tree (GBDT) method [19].

Deep CNNs have been widely used to classify and detect various plant diseases—biotic stress [112–115]. Image recognition and classification of maize leaf diseases, including northern corn leaf blight (*Exserohilum*), common rust (*Puccinia sorghi*), and gray leaf spot (*Cercospora*) diseases, have been conducted using deep CNN with an accuracy of 93.35% [116]. In cucumber (*Cucumis sativus*), a semantic segmentation model based on CNN was developed to segment the powdery mildew disease on leaf images at pixel level, and pixel accuracy of CNN model (96.08%) was more than segmentation methods of K-means, Random forest, and GBDT [28]. In pearl millet (*Pennisetum glaucum*), DNNs has been applied for identification of mildew disease, and accuracy of 95.00% was reported for the developed model [37].

### 2.5. Classical Mating Designs and Hybrid Breeding Programs

The integration of statistics into genetics led to some classical mating designs such as mean generation analysis [117], diallel crosses analysis [118–120], line  $\times$  tester analysis [121], North Carolina designs [122], and triple test cross [123,124]. These methods have been used for genetic analysis of crops in order to find the nature of gene actions (additive, dominance, and epistasis) involved in controlling important morphological, phenological, and yield component characteristics, to calculate broad and narrow sense heritability and predict the outcomes of cross-breeding programs.

The prediction of parental combinations is critical to the choice of superior combinational homozygous parental lines in F1-hybrid breeding programs [125]. However, it is a challenging task with a large number of cross combinations when there are many inbred parental lines. Therefore, the prediction of the yield performance of cross combinations of parental lines may significantly reduce the required time and budget of F1-hybrid breeding programs [126]. ANN could be used to predict parental combinations and calculate the correct values of general and specific combining abilities (GCA and SCA) in mating designs, such as topcross, line  $\times$  tester, and diallel cross. Khaki et al. [126] applied matrix factorization and a neural network to predict the yield performance of cross combinations of inbreds and testers of unsown maize on the basis of historical yield data collected from the crossing of other inbreds and testers. The proposed model was significantly better than other models such as deep factorization machines (DeepFM), generalized matrix factorization (GMF), LASSO, RF, and neural networks.

## 3. Applications of Machine Learning in In Vitro-Based Plant Biotechnology

Biotechnology-based breeding methods (BBBMs) complement classical breeding methods in rapid plant improvement. In vitro regeneration, as the main core of many in-vitro-based breeding methods, has numerous plant breeding applications. In situ and ex situ conservation and micropropagation (proliferation) are direct applications of in vitro regeneration [127]. In endangered rare plant species, like medicinal plants, in vitro culture is an effective strategy for mass propagation, germplasm

conservation, and production of bioactive compounds [128]. Several factors determine the fate of cultured cells in in vitro regeneration of plants. These are the plant genotype, plant growth regulators (PGRs), culture medium components, explant type, explant age, enhancer additives-elicitors, etc. [127]. These factors can be divided into three main categories: initial triggers of regeneration (environmental signal inputs and physical stimuli), epigenetic and transcriptional cellular responses to the initial triggers, and molecules that manage the formation and development of the new stem cell niche [129]. The combination and interactions between these factors lead to multifactorial nature of the in vitro plant regeneration process. Basal culture medium components, plant genotype, PGRs, explant type, and explant age are all multilevel factors with different applicable combinations. The inclusion of other factors results in a very complex situation for interpretation. Plant cells and tissues have nondeterministic and nonlinear developmental patterns in a stressful in vitro environment [130]. The analysis of variance of factorial experiments and simple means comparison analysis with classical methods such as LSD, Tukey's HSD, and Duncan's test, are the main statistical methods used to interpret the effects of interaction between effective factors in most in vitro regeneration studies [128,131,132].

Murashige and Skoog (MS), modified MS (MMS), Gamborg's B5 medium Woody Plant Medium (WPM), and Driver and Kuniyuki Woody Plant Medium (DKW) are the most commonly used basal culture media in in vitro regeneration studies. Basal medium manipulation is a promoting strategy that has been applied to increase the output of in vitro studies [133]. However, due to the large number of micro- and macroelements in the culture medium, it is difficult to manipulate their concentrations. In this situation, prediction of the effect of culture media components on the target characteristics of in vitro regenerants is the right solution. Artificial neural networks have been applied in these experiments to predict the best culture media components for efficient propagation of different plant species [29,31,134].

Different combinations of auxin and cytokinin PGRs can determine the developmental fate of cultured cells and tissues toward organogenesis and/or somatic embryogenesis. The cytokinin/auxin ratio is also very important in in vitro studies [135]. Niazian et al. [131] found that 2,4-dichlorophenoxyacetic acid (2,4-D) combined with kinetin resulted in indirect somatic embryogenesis of cultured hypocotyl segments of ajowan medicinal plants, whereas a combination of 3-methoxy(-6-benzylamino-9-tetrahydropyran-2-yl) purine and naphthalene acetic acid led cultivated explants toward an indirect shoot regeneration pathway. Arab et al. [30] combined artificial neural networks and genetic algorithms to predict and optimize the effect of cytokinin–auxin plant hormone (BAP, KIN, TDZ, IBA, and NAA) combinations and concentrations on the number of microshoots per explant, the length of microshoots, developed callus weight, and the quality index of plantlets in in vitro proliferation of Garnem (G × N15) rootstock. The ANN model predicted the number and length of microshoots with high accuracy. The highest values of the variable sensitivity ratio for the proliferation rate were related to the BAP (19.3), KIN (9.64), and IBA (2.63) inputs. An MLP-ANN was developed to predict the physical properties of embryogenic callus and the number of somatic embryos in in vitro regeneration of ajowan under the effect of different combinations of the explant age, concentrations of 2,4-D, kinetin, and sucrose inputs [25]. The ANN model predicted the physical properties of embryogenic callus (area, perimeter, Feret diameter, roundness, and true density) and the number of somatic embryos better than the multiple linear regressions. Fifteen-day-old hypocotyl explants × 1.5 mg/L 2,4-D × 0.5 mg/L Kin × 2.5% (*w/v*) sucrose was the best combination of inputs with the highest measured and predicted number of somatic embryos [25].

Apart from culture medium components and PGRs combination, ANN has been applied to model the sterilization step of in vitro regeneration. Hesami et al. [27] applied an MLP-ANN along with a genetic algorithm to model and optimize the contamination frequency and explant viability under the influence of seven input variables, i.e., HgCl<sub>2</sub>, Ca(ClO)<sub>2</sub>, nanosilver, H<sub>2</sub>O<sub>2</sub>, NaOCl, AgNO<sub>3</sub>, and immersion times, in an in vitro culture of chrysanthemum. The lowest contamination frequency (0%) and the highest explant viability (99.98%) resulted from 1.62% NaOCl at 13.96 min immersion time. The sensitivity analysis of the ANN showed that the immersion time was the most important

variable affecting the contamination frequency and explant viability [27]. ANNs are also used to simulate in vitro growth of plant tissue cultures, distinguish embryos from nonembryos, predict the formation of plantlets from embryos, estimate the biomass of cell cultures, simulate the distribution of temperature in a culture vessel, identify and estimate the in vitro induced shoot length, and cluster in vitro regenerated plantlets [130].

Other in vitro-based breeding methods, such as artificial polyploidy induction, doubled haploid production, plant gene transformation, and genome editing methods also have multifactorial nature and require multivariate statistical methods to interpret the results. Different chemical enhancers can be used in in vitro doubled haploid production methods (induced parthenogenesis and androgenesis) to improve the haploid induction efficiency, e.g., PGRs, osmoprotectants, cellular antioxidants, reactive oxygen species scavengers, polyamines, stress hormones, chlormequat chloride, compatible solutes, DNA demethylating agents, histone deacetylase inhibitors, cell wall remodeling agents, ethylene inhibitors, and other applicable additives. They enhance tolerance to inductive stresses and improve the final efficiency of doubled haploid production [108]. ANN models may improve the efficiency of in vitro doubled haploid production and solve the problem of recalcitrant species/genotypes by predicting the best combination(s) of these additives in interaction with other influencing factors, such as the plant genotype, the surrounding environment of donor plants, physical treatments (inductive stresses) of cultured gametophytic cells, the developmental stage of initial gametophytic cells, and culture medium components. The ANN predicted the callus induction percentage in androgenesis (anther culture) of tomato (*Lycopersicon esculentum* L.) under the influence of plant genotype, the concentrations of 2,4-D and kinetin PGRs, and the concentration of gum Arabic better than the MLR model [50].

Plants' vigor and performance are commonly enhanced by mitotic-induced polyploidy. It consists in in vivo and in vitro application of mitotic spindle poisons [136]. In vitro-induced polyploidy is a multifactorial procedure. The efficiency of in vitro-induced polyploidy may be affected not only by in vitro regeneration parameters (basal culture medium components, combination of PGRs, additives, etc.) but also by the plant genotype, the developmental stage of initial explants as well as the type, dosage, and duration (exposure time) of the application of the antimitotic agent. Due to the genotype dependency, different genotypes of plant species exhibit different responses to concentrations of the antimitotic agent applied [137]. This results in significant interaction of the plant genotype and antimitotic agent in artificial polyploidy induction. Although there have been no reports on the application of ANN to model and predict the results of in vitro-induced artificial polyploidy, it might increase the efficiency by predicting and finding the best combination and interaction of all influential factors.

*Agrobacterium*-mediated gene transformation is a well-known method of plant gene transformation and genetic engineering. However, various parameters must be optimized for an efficient gene delivery, including the *Agrobacterium* strain cell density, the time of inoculation, the type and concentration of antibiotics to kill *Agrobacterium*, the type and concentration of selectable antibiotics, and the concentration of acetosyringone [138]. These influencing factors along with in vitro regeneration factors result in a multi-variable nature of *Agrobacterium*-mediated gene transformation [127]. It is obvious that machine learning algorithms could be used to predict and optimize *Agrobacterium*-mediated gene transformation, especially in important *Agrobacterium*-recalcitrant plant species.

#### 4. Coupled Machine Learning-Image Processing for High-Throughput Phenotyping and Precision Agriculture

Classical measurement of plants' physical features by visual assessment is a laborious, time-consuming, costly, and error-prone process in both conventional and in vitro-based plant breeding studies. This step can be accelerated and facilitated by the machine vision method, which is more accurate and precise than visual assessment. Nondestructive measurement of physical features, both outdoors and in vitro, is another important advantage of image processing [25]. Automated non-invasive fast scoring of several plant traits through high-throughput phenotyping platforms can



speed up and facilitate the phenotyping of plant populations and selection of superior varieties [139]. The integration of precise measured image-based characteristics with omics data could help to identify the key traits involved in the mechanisms of stress tolerance and acclimation [109]. On the other hand, the ability of deep learning in the identification of plants' features provides a great opportunity for further advances in image analysis [98]. Combined image processing (for feature extraction) and machine learning (for data analysis) is a powerful strategy required for faster and precise image-based plant phenotyping [140]. The use of deep learning techniques in computer vision can accelerate plant breeding programs such as plant phenotyping and classification of genotypes [141]. Coupled image processing-ANN has been used to measure phenotypic characteristics and assess genetic diversity and classify different plant species [38,65,142,143]. Deep learning, especially CNN, has become a powerful tool for image analysis in recent years [49]. Uzal et al. [48] applied a computer vision method for feature extraction along with developed convolutional neural networks to estimate the number of seeds in soybean pods and then to classify the obtained data. In most cases, the convolutional neural networks learnt to detect each seed in the pod, which indicates their high classification efficiency. There are other advanced imaging techniques, which are more efficient than simple visualization techniques and can be used to analyze in-field images instead of indoor methods. Recently, an R-based pipeline has been developed, which enables analysis of orthomosaic images from agricultural field trials and calculation of the number of plants per plot, canopy cover percentage, vegetation indices, and plant height [139]. A deep neural network model trained with such in-field images could very effectively classify and estimate desired characteristics from in-field images [48]. Coupled image processing-artificial neural network has been used in BBMs for in vitro modeling of somatic embryogenesis in ajowan [25] and androgenesis-based haploid induction in tomato [50]. Plant phenotyping and precision agriculture could be significantly different in terms of the spatial and temporal resolutions, although both generate big data sets in a format of image. These are information- and technology-based domains with specific demand and challenges. Precision agriculture is an agricultural management system based on spatial and temporal variability in crop and soil factors within a field (with environmental parameters). However, in phenotyping systems, the crop field parameters are homogenous and datasets in molecular, cellular, and whole plant levels are considered for plant phenotyping. Precision agriculture examines spatial heterogeneities within crop stands, whereas the appearance and performance of a genotype under distinct environmental conditions are examined in plant phenotyping [144,145]. High-throughput salt-stress phenotyping has been reported in okra (*Abelmoschus esculentus* L.) through a trained DNN using physiological and biochemical traits, such as fresh weight, SPAD, elemental contents, and photosynthesis-related parameters, measured from 13 genotypes under salt stress treatment [36]. Establishment of high-throughput phenotyping platforms (HTPPs) to phenotype physiomorphological traits under highly heterogeneous field environment, in a precise, labor-, and cost-effective manner, is essential to bridge the gap between genomics and phenomics [146]. Machine learning algorithms can be used for image-based plant stress phenotyping in a wide scale from leaf and canopy to field range. Identification, classification, quantification, and prediction of big data, obtained from higher-throughput phenotyping systems such as unmanned aerial system (UAS) technology and ground robots, can be conducted through deep learning algorithms [147]. In carrot (*Daucus carota*), a precision agriculture approach was conducted through on-farm punctual carrot sampling data incorporated into the satellite imagery data using a random forest regression algorithm. Accuracy of developed model to predict carrot yield using database composed of spectral bands was acceptable ( $R^2 = 0.82$ ; RMSE = 2.64 mg ha<sup>-1</sup>; MAE = 1.74 mg ha<sup>-1</sup>) [26].

## 5. A Proposed Idea for Plant Ploidy Level Determination through Image Processing-Machine Learning

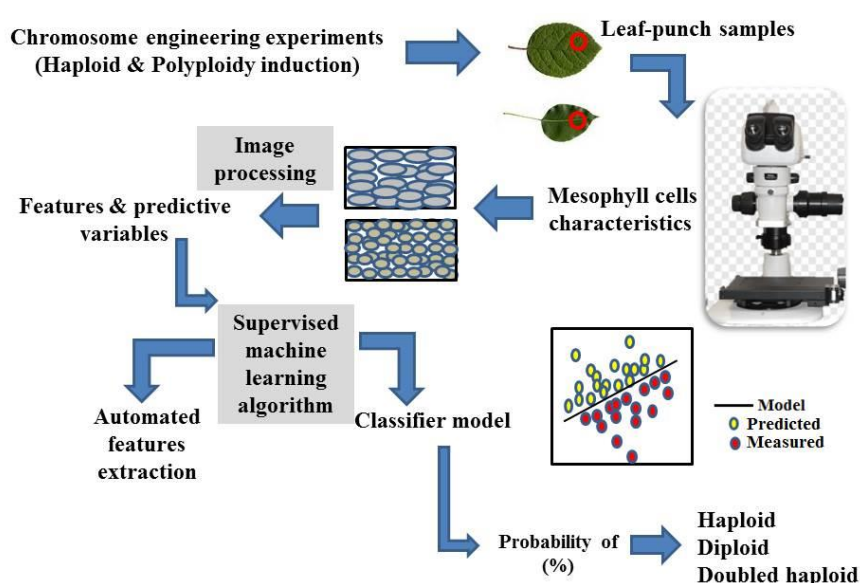
In chromosome engineering studies (polyploidy and haploid induction), one important step is taken to verify the ploidy level. It can be confirmed through direct (chromosome counting) or indirect methods (morphological and anatomical indicators and flow cytometry). Although the direct



method of chromosome counting is reliable and unambiguous [148], it is laborious, time consuming, and complicated and requires highly skilled operators [149]. Indirect verification of the ploidy level through classical markers, such as stomatal morphometric data (stomatal density per unit area, the number and size of stomata), the density of chloroplasts per stomatal guard cells, size of guard cells and pollen size, is a rapid and simple method [150], but not completely reliable. Flow cytometry is a reliable method based on direct correlation between the nuclear DNA content and ploidy level. However, according to a recent study, the comparison of the DNA content in standardized leaf punch samples is not a reliable method to recognize putative doubled haploids, as there is a DNA content equivalence between haploid and diploid samples [151].

Machine learning algorithms can be used for ploidy level identification of plants. Recently, a deep learning-based object detection algorithm has been developed for evaluating the stomatal density and elucidating the variation in the stomatal density among various soybean accessions [49]. This DNN could also be useful for ploidy level prediction. There have been two reports on the use of other methods to identify the ploidy level in plants. Altuntaş et al. [33] used convolutional neural networks to recognize haploid and diploid maize seeds through *R1-nj* anthocyanin color marker data of 1230 haploid and 1770 diploid maize seed images. The accuracy and sensitivity of the model amounted to 94.22% and 94.58%, respectively [33]. Remote sensing has also been applied to determine the ploidy level of quaking aspen (*Populus tremuloides* Michx.) [152].

Here, we offer another idea to identify the ploidy level of plants through coupled image processing-supervised deep neural network using visual data of cellular patterning of the epidermal layer. Haploids have smaller and more densely packed epidermal and mesophyll cells (more cells per same unit area) than doubled haploids. This results in an equivalent DNA content per unit leaf area for haploids and their counterpart diploids [151]. Cellular patterning in the epidermis and mesophyll can be specific to each ploidy group. Therefore, epidermal cell patterning (size shape and number) could be used as ploidy level recognition and classification criteria [151]. The use of imaging techniques for precise feature extraction of leaf punch samples (the cellular pattern, including the cell size and number) and the subsequent modeling of captured images (classification modeling) through deep learning approaches, particularly CNN, results in an image-based model, which can be used to estimate the ploidy level in chromosome engineering studies of different plant species (Figure 3). It is a more precise, fast, and cost-effective method of ploidy level distinction, which could also be used in other branches of plant science, e.g., in genetic diversity, evolutionary, and species invasiveness studies.



**Figure 3.** The proposed coupled image processing-supervised machine learning to determine the ploidy level in different plant species through cellular patterning in leaf epidermis and mesophyll layers.

## 6. Conclusions

Most classical statistical methods use only simple statistics and few influential factors to assess the biological features of plants. For example,  $Y_p$  and  $Y_s$  are the only indices used to identify drought-tolerant plant genotypes in yield-based drought tolerance assessment methods. However, there are other influential factors, such as cellular, physiological, and phytochemical pathways, which are involved in plants' responses to environmental stress. The tolerance of different plant species to biotic and abiotic stresses, as complex biological processes, can be efficiently enhanced through large-scale analysis of phenomic, metabolomic, and genomic data. Machine learning models are capable of processing large amounts of data (imaging and remote-sensing data) for high-throughput stress phenotyping. The analysis of different omics and phenomic data may result in more precise interpretation of GEI and yield stability. Plants' qualitative and quantitative characteristics can be predicted more precisely by analysis of climate data (temperature, humidity, sunshine, precipitation, etc.), soil factors, agricultural operations data (harvest date, information on diseases, crop status, ground temperature, etc.), topographic, and meteorological data. Big data analysis enables more efficient classification of plants' phenotypes and genotypes. Machine learning techniques are able to manage large amounts of data in various areas of plant breeding, which can lead to more accurate results and better interoperation than classical statistical methods. Artificial neural networks can be used for pattern recognition, nonlinear regression, and classification purposes in plant tissue culture studies because they can handle binary, continuous, categorical, and fuzzy datasets. The present review can give an overview of applications of machine learning to plant breeders. It would be helpful to adopt the correct method of data analysis in future studies, which in turn can increase the output of studies.

**Author Contributions:** M.N. and G.N. contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** The publication was co-financed within the framework of the Ministry of Science and Higher Education program titled "Regional Initiative Excellence" in 2019–2022, project no. 005/RID/2018/19.

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

## References

1. Ali, S.; Shafique, O.; Mahmood, T.; Hanif, M.A.; Ahmed, I.; Khan, B.A. A Review about Perspectives of Nanotechnology in Agriculture. *Pakistan J. Agric. Res.* **2018**, *31*, 116–121. [\[CrossRef\]](#)
2. Niazian, M.; Sadat-Noori, S.A.; Abdipour, M. Modeling the seed yield of Ajowan (*Trachyspermum ammi* L.) using artificial neural network and multiple linear regression models. *Ind. Crops Prod.* **2018**, *117*, 224–234. [\[CrossRef\]](#)
3. Hesami, M.; Naderi, R.; Tohidfar, M.; Yoosefzadeh-Najafabadi, M. Application of Adaptive Neuro-Fuzzy Inference System-Non-dominated Sorting Genetic Algorithm-II (ANFIS-NSGAI) for Modeling and Optimizing Somatic Embryogenesis of Chrysanthemum. *Front. Plant Sci.* **2019**, *10*. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Chegini, G.R.; Khazaei, J.; Ghobadian, B.; Goudarzi, A.M. Prediction of process and product parameters in an orange juice spray dryer using artificial neural networks. *J. Food Eng.* **2008**, *84*, 534–543. [\[CrossRef\]](#)
5. Zheng, H.; Li, W.; Jiang, J.; Liu, Y.; Cheng, T.; Tian, Y.; Zhu, Y.; Cao, W.; Zhang, Y.; Yao, X.A. Comparative Assessment of Different Modeling Algorithms for Estimating Leaf Nitrogen Content in Winter Wheat Using Multispectral Images from an Unmanned Aerial Vehicle. *Remote Sens.* **2018**, *10*, 2026. [\[CrossRef\]](#)
6. Hesami, M.; Naderi, R.; Yoosefzadeh-Najafabadi, M.; Rahmati, M. Data-driven modeling in plant tissue culture. *J. Appl. Environ. Biol. Sci.* **2017**, *7*, 37–44.
7. Salehi, M.; Farhadi, S.; Moieni, A.; Safaie, N.; Ahmadi, H. Mathematical Modeling of Growth and Paclitaxel Biosynthesis in *Corylus avellana* Cell Culture Responding to Fungal Elicitors Using Multilayer Perceptron-Genetic Algorithm. *Front. Plant Sci.* **2020**, *11*, 1148. [\[CrossRef\]](#)
8. Asefpour Vakilian, K. Machine learning improves our knowledge about miRNA functions towards plant abiotic stresses. *Sci. Rep.* **2020**, *10*, 3041. [\[CrossRef\]](#)

9. Wang, H.; Cimen, E.; Singh, N.; Buckler, E. Deep learning for plant genomics and crop improvement. *Curr. Opin. Plant Biol.* **2020**, *54*, 34–41. [\[CrossRef\]](#)
10. Hu, H.; Scheben, A.; Edwards, D. Advances in Integrating Genomics and Bioinformatics in the Plant Breeding Pipeline. *Agriculture* **2018**, *8*, 75. [\[CrossRef\]](#)
11. Orozco-Arias, S.; Isaza, G.; Guyot, R. Retrotransposons in Plant Genomes: Structure, Identification, and Classification through Bioinformatics and Machine Learning. *Int. J. Mol. Sci.* **2019**, *20*, 3837. [\[CrossRef\]](#)
12. 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. [\[CrossRef\]](#)
13. Azevedo, A.M.; Andrade Júnior, V.C.D.; Pedrosa, C.E.; Oliveira, C.M.D.; Dornas, M.F.S.; Cruz, C.D.; Valadares, N.R. Application of artificial neural networks in indirect selection: A case study on the breeding of lettuce. *Bragantia* **2015**, *74*, 387–393. [\[CrossRef\]](#)
14. Gu, J.; Wang, Z.; Kuen, J.; Ma, L.; Shahroudy, A.; Shuai, B.; Liu, T.; Wang, X.; Wang, G.; Cai, J.; et al. Recent advances in convolutional neural networks. *Pattern Recognit.* **2018**, *77*, 354–377. [\[CrossRef\]](#)
15. Hesami, M.; Naderi, R.; Tohidfar, M.; Yoosefzadeh-Najafabadi, M. Development of support vector machine-based model and comparative analysis with artificial neural network for modeling the plant tissue culture procedures: Effect of plant growth regulators on somatic embryogenesis of chrysanthemum, as a case study. *Plant Methods* **2020**, *16*, 112. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Ali, A.M.; Darvishzadeh, R.; Skidmore, A.; Gara, T.W.; Heurich, M. Machine learning methods' performance in radiative transfer model inversion to retrieve plant traits from Sentinel-2 data of a mixed mountain forest. *Int. J. Digit. Earth* **2020**, 1–15. [\[CrossRef\]](#)
17. Gold, K.M.; Townsend, P.A.; Herrmann, I.; Gevens, A.J. Investigating potato late blight physiological differences across potato cultivars with spectroscopy and machine learning. *Plant Sci.* **2020**, *295*, 110316. [\[CrossRef\]](#)
18. Barth, R.; IJsselmuiden, J.; Hemming, J.; Van Henten, E.J. Synthetic bootstrapping of convolutional neural networks for semantic plant part segmentation. *Comput. Electron. Agric.* **2019**, *161*, 291–304. [\[CrossRef\]](#)
19. An, J.; Li, W.; Li, M.; Cui, S.; Yue, H. Identification and Classification of Maize Drought Stress Using Deep Convolutional Neural Network. *Symmetry* **2019**, *11*, 256. [\[CrossRef\]](#)
20. Geetharamani, G.; Arun Pandian, J. Identification of plant leaf diseases using a nine-layer deep convolutional neural network. *Comput. Electr. Eng.* **2019**, *76*, 323–338.
21. Kattenborn, T.; Eichel, J.; Wiser, S.; Burrows, L.; Fassnacht, F.E.; Schmidtlein, S. Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. *Remote Sens. Ecol. Conserv.* **2020**. [\[CrossRef\]](#)
22. Iniyan, S.; Jebakumar, R.; Mangalraj, P.; Mohit, M.; Nanda, A. Plant Disease Identification and Detection Using Support Vector Machines and Artificial Neural Networks. In *Artificial Intelligence and Evolutionary Computations in Engineering Systems; Advances in Intelligent Systems and Computing*; Dash, S., Lakshmi, C., Das, S., Panigrahi, B., Eds.; Springer: Singapore, 2020; pp. 15–27.
23. Wang, Y.; Li, T.; Jin, G.; Wei, Y.; Li, L.; Kalkhaje, Y.K.; Ning, J.; Zhang, Z. Qualitative and quantitative diagnosis of nitrogen nutrition of tea plants under field condition using hyperspectral imaging coupled with chemometrics. *J. Sci. Food Agric.* **2020**, *100*, 161–167. [\[CrossRef\]](#)
24. Niazi, M.; Sadat-Noori, S.A.; Abdipour, M. Artificial neural network and multiple regression analysis models to predict essential oil content of ajowan (*Carum copticum* L.). *J. Appl. Res. Med. Aromat. Plants* **2018**, *9*, 124–131. [\[CrossRef\]](#)
25. Niazi, M.; Sadat-Noori, S.A.; Abdipour, M.; Tohidfar, M.; Mortazavian, S.M.M. Image Processing and Artificial Neural Network-Based Models to Measure and Predict Physical Properties of Embryogenic Callus and Number of Somatic Embryos in Ajowan (*Trachyspermum ammi* (L.) Sprague). *Vitr. Cell. Dev. Biol. Plant* **2018**, *54*, 54–68. [\[CrossRef\]](#)
26. Wei, M.C.F.; Maldaner, L.F.; Ottoni, P.M.N.; Molin, J.P. Carrot Yield Mapping: A Precision Agriculture Approach Based on Machine Learning. *AI* **2020**, *1*, 229–241.
27. Hesami, M.; Naderi, R.; Tohidfar, M. Modeling and Optimizing in vitro Sterilization of Chrysanthemum via Multilayer Perceptron-Non-dominated Sorting Genetic Algorithm-II (MLP-NSGAI). *Front. Plant Sci.* **2019**, *10*. [\[CrossRef\]](#) [\[PubMed\]](#)
28. Lin, K.; Gong, L.; Huang, Y.; Liu, C.; Pan, J. Deep Learning-Based Segmentation and Quantification of Cucumber Powdery Mildew Using Convolutional Neural Network. *Front. Plant Sci.* **2019**, *10*. [\[CrossRef\]](#)

29. Arab, M.M.; Yadollahi, A.; Shojaeiyan, A.; Ahmadi, H. Artificial Neural Network Genetic Algorithm as Powerful Tool to Predict and Optimize In vitro Proliferation Mineral Medium for G × N15 Rootstock. *Front. Plant Sci.* **2016**, *7*. [[CrossRef](#)]
30. Arab, M.M.; Yadollahi, A.; Ahmadi, H.; Eftekhari, M.; Maleki, M. Mathematical Modeling and Optimizing of in Vitro Hormonal Combination for G × N15 Vegetative Rootstock Proliferation Using Artificial Neural Network-Genetic Algorithm (ANN-GA). *Front. Plant Sci.* **2017**, *8*. [[CrossRef](#)]
31. Arab, M.M.; Yadollahi, A.; Eftekhari, M.; Ahmadi, H.; Akbari, M.; Khorami, S.S. Modeling and Optimizing a New Culture Medium for In Vitro Rooting of G×N15 Prunus Rootstock using Artificial Neural Network-Genetic Algorithm. *Sci. Rep.* **2018**, *8*, 9977. [[CrossRef](#)]
32. Costa, M.O.; Capel, L.S.; Maldonado, C.; Mora, F.; Mangolin, C.A.; Machado, M. High genetic differentiation of grapevine rootstock varieties determined by molecular markers and artificial neural networks. *Acta Sci. Agron.* **2019**, *42*, e43475. [[CrossRef](#)]
33. Altuntaş, Y.; Cömert, Z.; Kocamaz, A.F. Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. *Comput. Electron. Agric.* **2019**, *163*, 104874. [[CrossRef](#)]
34. Darwish, A.; Ezzat, D.; Hassanien, A.E. An optimized model based on convolutional neural networks and orthogonal learning particle swarm optimization algorithm for plant diseases diagnosis. *Swarm Evol. Comput.* **2020**, *52*, 100616. [[CrossRef](#)]
35. Mishra, S.; Sachan, R.; Rajpal, D. Deep Convolutional Neural Network based Detection System for Real-time Corn Plant Disease Recognition. *Procedia Comput. Sci.* **2020**, *167*, 2003–2010. [[CrossRef](#)]
36. Feng, X.; Zhan, Y.; Wang, Q.; Yang, X.; Yu, C.; Wang, H.; Tang, Z.; Jiang, D.; Peng, C.; He, Y. Hyperspectral imaging combined with machine learning as a tool to obtain high-throughput plant salt-stress phenotyping. *Plant J.* **2020**, *101*, 1448–1461. [[CrossRef](#)]
37. Coulibaly, S.; Kamsu-Foguem, B.; Kamissoko, D.; Traore, D. Deep neural networks with transfer learning in millet crop images. *Comput. Ind.* **2019**, *108*, 115–120. [[CrossRef](#)]
38. Azizi, A.; Abbaspour-Gilandeh, Y.; Nooshyar, M.; Afkari-Sayah, A. Identifying Potato Varieties Using Machine Vision and Artificial Neural Networks. *Int. J. Food Prop.* **2016**, *19*, 618–635. [[CrossRef](#)]
39. Niedbała, G.; Piekutowska, M.; Weres, J.; Korzeniewicz, R.; Witaszek, K.; Adamski, M.; Pilarski, K.; Czechowska-Kosacka, A.; Krysztofiak-Kaniewska, A. Application of Artificial Neural Networks for Yield Modeling of Winter Rapeseed Based on Combined Quantitative and Qualitative Data. *Agronomy* **2019**, *9*, 781. [[CrossRef](#)]
40. Zhang, J.; Zhao, B.; Yang, C.; Shi, Y.; Liao, Q.; Zhou, G.; Wang, C.; Xie, T.; Jiang, Z.; Zhang, D.; et al. Rapeseed Stand Count Estimation at Leaf Development Stages With UAV Imagery and Convolutional Neural Networks. *Front. Plant Sci.* **2020**, *11*. [[CrossRef](#)]
41. Niedbała, G. Application of artificial neural networks for multi-criteria yield prediction of winter rapeseed. *Sustainability* **2019**, *11*, 533.
42. Niedbała, G. Simple model based on artificial neural network for early prediction and simulation winter rapeseed yield. *J. Integr. Agric.* **2019**, *18*, 54–61. [[CrossRef](#)]
43. Li, D.; Wang, R.; Xie, C.; Liu, L.; Zhang, J.; Li, R.; Wang, F.; Zhou, M.; Liu, W. A Recognition Method for Rice Plant Diseases and Pests Video Detection Based on Deep Convolutional Neural Network. *Sensors* **2020**, *20*, 578. [[CrossRef](#)] [[PubMed](#)]
44. Rahman, C.R.; Arko, P.S.; Ali, M.E.; Iqbal Khan, M.A.; Apon, S.H.; Nowrin, F.; Wasif, A. Identification and recognition of rice diseases and pests using convolutional neural networks. *Biosyst. Eng.* **2020**, *194*, 112–120. [[CrossRef](#)]
45. Abdipour, M.; Younessi-Hmazekhanlu, M.; Ramazani, S.H.R.; Omid, A. Hassan 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](#)]
46. 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](#)]
47. Parsaeian, M.; Shahabi, M.; Hassanpour, H. Estimating Oil and Protein Content of Sesame Seeds Using Image Processing and Artificial Neural Network. *J. Am. Oil Chem. Soc.* **2020**, *97*, 691–702. [[CrossRef](#)]
48. Uzal, L.C.; Grinblat, G.L.; Namías, R.; Larese, M.G.; Bianchi, J.S.; Morandi, E.N.; Granitto, P.M. Seed-per-pod estimation for plant breeding using deep learning. *Comput. Electron. Agric.* **2018**, *150*, 196–204. [[CrossRef](#)]



49. Sakoda, K.; Watanabe, T.; Sukemura, S.; Kobayashi, S.; Nagasaki, Y.; Tanaka, Y.; Shiraiwa, T. Genetic Diversity in Stomatal Density among Soybeans Elucidated Using High-throughput Technique Based on an Algorithm for Object Detection. *Sci. Rep.* **2019**, *9*, 7610. [\[CrossRef\]](#)
50. Niazian, M.; Shariatpanahi, M.E.; Abdipour, M.; Oroojloo, M. Modeling callus induction and regeneration in an anther culture of tomato (*Lycopersicon esculentum* L.) using image processing and artificial neural network method. *Protoplasma* **2019**, *256*, 1317–1332. [\[CrossRef\]](#)
51. Verma, S.; Chug, A.; Singh, A.P. Application of convolutional neural networks for evaluation of disease severity in tomato plant. *J. Discret. Math. Sci. Cryptogr.* **2020**, *23*, 273–282. [\[CrossRef\]](#)
52. Ravari, S.Z.; Dehghani, H.; Naghavi, H. Assessment of salinity indices to identify Iranian wheat varieties using an artificial neural network. *Ann. Appl. Biol.* **2016**, *168*, 185–194. [\[CrossRef\]](#)
53. Niedbała, G.; Kozłowski, R.J. Application of Artificial Neural Networks for Multi-Criteria Yield Prediction of Winter Wheat. *J. Agric. Sci. Technol.* **2019**, *21*, 51–61.
54. Niedbała, G.; Nowakowski, K.; Rudowicz-Nawrocka, J.; Piekutowska, M.; Weres, J.; Tomczak, R.J.; Tyksiński, T.; Pinto, A.Á. Multicriteria prediction and simulation of winter wheat yield using extended qualitative and quantitative data based on artificial neural networks. *Appl. Sci.* **2019**, *9*, 2773. [\[CrossRef\]](#)
55. Sadeghi-Tehran, P.; Virlet, N.; Ampe, E.M.; Reyns, P.; Hawkesford, M.J. DeepCount: In-Field Automatic Quantification of Wheat Spikes Using Simple Linear Iterative Clustering and Deep Convolutional Neural Networks. *Front. Plant Sci.* **2019**, *10*. [\[CrossRef\]](#) [\[PubMed\]](#)
56. Haider, S.A.; Naqvi, S.R.; Akram, T.; Umar, G.A.; Shahzad, A.; Sial, M.R.; Khaliq, S.; Kamran, M. LSTM Neural Network Based Forecasting Model for Wheat Production in Pakistan. *Agronomy* **2019**, *9*, 72. [\[CrossRef\]](#)
57. Ma, W.; Qiu, Z.; Song, J.; Li, J.; Cheng, Q.; Zhai, J.; Ma, C. A deep convolutional neural network approach for predicting phenotypes from genotypes. *Planta* **2018**, *248*, 1307–1318. [\[CrossRef\]](#)
58. Hesami, M.; Condori-Apfata, J.A.; Valderrama Valencia, M.; Mohammadi, M. Application of Artificial Neural Network for Modeling and Studying In Vitro Genotype-Independent Shoot Regeneration in Wheat. *Appl. Sci.* **2020**, *10*, 5370. [\[CrossRef\]](#)
59. Niedbała, G.; Kurasiak-Popowska, D.; Stuper-Szablewska, K.; Nawracała, J. Application of Artificial Neural Networks to Analyze the Concentration of Ferulic Acid, Deoxynivalenol, and Nivalenol in Winter Wheat Grain. *Agriculture* **2020**, *10*, 127. [\[CrossRef\]](#)
60. Ray, A.; Halder, T.; Jena, S.; Sahoo, A.; Ghosh, B.; Mohanty, S.; Mahapatra, N.; Nayak, S. Application of artificial neural network (ANN) model for prediction and optimization of coronarin D content in Hedychium coronarium. *Ind. Crops Prod.* **2020**, *146*, 112186. [\[CrossRef\]](#)
61. Srivastava, A.; Gupta, S.; Shanker, K.; Gupta, N.; Gupta, A.K.; Lal, R.K. Genetic diversity in Indian poppy (*P. somniferum* L.) germplasm using multivariate and SCoT marker analyses. *Ind. Crops Prod.* **2020**, *144*, 112050. [\[CrossRef\]](#)
62. Niazian, M.; Sadat Noori, S.A.; Tohidfar, M.; Mortazavian, S.M.M. Essential Oil Yield and Agro-morphological Traits in Some Iranian Ecotypes of Ajowan (*Carum copticum* L.). *J. Essent. Oil Bear. Plants* **2017**, *20*, 1151–1156. [\[CrossRef\]](#)
63. Schulman, A.H. Molecular markers to assess genetic diversity. *Euphytica* **2007**, *158*, 313–321. [\[CrossRef\]](#)
64. Boonsrangsom, T. Genetic diversity of ‘Wan Chak Motluk’ (*Curcuma comosa* Roxb.) in Thailand using morphological characteristics and random amplification of polymorphic DNA (RAPD) markers. *South Afr. J. Bot.* **2020**, *130*, 224–230. [\[CrossRef\]](#)
65. Pandolfi, C.; Mugnai, S.; Azzarello, E.; Bergamasco, S.; Masi, E.; Mancuso, S. Artificial neural networks as a tool for plant identification: A case study on Vietnamese tea accessions. *Euphytica* **2009**, *166*, 411–421. [\[CrossRef\]](#)
66. Raza, A.; Mehmood, S.S.; Ashraf, F.; Khan, R.S.A. Genetic Diversity Analysis of Brassica Species Using PCR-Based SSR Markers. *Gesunde Pflanz.* **2019**, *71*, 1–7. [\[CrossRef\]](#)
67. Bird, C.; Schweizer, M.; Roberts, A.; Austin, W.E.N.; Knudsen, K.L.; Evans, K.M.; Filipsson, H.L.; Sayer, M.D.J.; Geslin, E.; Darling, K.F. The genetic diversity, morphology, biogeography, and taxonomic designations of Ammonia (Foraminifera) in the Northeast Atlantic. *Mar. Micropaleontol.* **2020**, *155*, 101726. [\[CrossRef\]](#)
68. Poletto, T.; Poletto, I.; Moraes Silva, L.M.; Brião Muniz, M.F.; Silveira Reiniger, L.R.; Richards, N.; Stefenon, V.M. Morphological, chemical and genetic analysis of southern Brazilian pecan (*Carya illinoensis*) accessions. *Sci. Hortic. Amst.* **2020**, *261*, 108863. [\[CrossRef\]](#)

69. Saini, G.; Khamparia, A.; Luhach, A.K. Classification of Plants Using Convolutional Neural Network. In *Advances in Intelligent Systems and Computing, Proceedings of the First International Conference on Sustainable Technologies for Computational Intelligence*; Luhach, A., Kosa, J., Poonia, R., Gao, X., Singh, D., Eds.; Springer: Singapore, 2020; Volume 1045, pp. 551–561.
70. Yang, H.-W.; Hsu, H.-C.; Yang, C.-K.; Tsai, M.-J.; Kuo, Y.-F. Differentiating between morphologically similar species in genus *Cinnamomum* (Lauraceae) using deep convolutional neural networks. *Comput. Electron. Agric.* **2019**, *162*, 739–748. [[CrossRef](#)]
71. Sant’Anna, I.C.; Tomaz, R.S.; Silva, G.N.; Nascimento, M.; Bhering, L.L.; Cruz, C.D. Superiority of artificial neural networks for a genetic classification procedure. *Genet. Mol. Res.* **2015**, *14*, 9898–9906. [[CrossRef](#)]
72. Korani, W.; Clevenger, J.P.; Chu, Y.; Ozias-Akins, P. Machine Learning as an Effective Method for Identifying True Single Nucleotide Polymorphisms in Polyploid Plants. *Plant Genome* **2019**, *12*, 1–10. [[CrossRef](#)]
73. González-Camacho, J.M.; de los Campos, G.; Pérez, P.; Gianola, D.; Cairns, J.E.; Mahuku, G.; Babu, R.; Crossa, J. Genome-enabled prediction of genetic values using radial basis function neural networks. *Theor. Appl. Genet.* **2012**, *125*, 759–771. [[CrossRef](#)] [[PubMed](#)]
74. Peixoto, L.A.; Bhering, L.L.; Cruz, C.D. Artificial neural networks reveal efficiency in genetic value prediction. *Genet. Mol. Res.* **2015**, *14*, 6796–6807. [[CrossRef](#)]
75. Zingaretti, L.M.; Gezan, S.A.; Ferrão, L.F.V.; Osorio, L.F.; Monfort, A.; Muñoz, P.R.; Whitaker, V.M.; Pérez-Enciso, M. Exploring Deep Learning for Complex Trait Genomic Prediction in Polyploid Outcrossing Species. *Front. Plant Sci.* **2020**, *11*. [[CrossRef](#)]
76. Ghaffari, H.; Tadayon, M.R.; Razmjoo, J.; Bahador, M.; Soureshjani, H.K.; Yuan, T. Impact of Jasmonic Acid on Sugar Yield and Physiological Traits of Sugar Beet in Response to Water Deficit Regimes: Using Stepwise Regression Approach. *Russ. J. Plant Physiol.* **2020**, *67*, 482–493. [[CrossRef](#)]
77. Zou, J.; Hu, W.; Li, Y.; He, J.; Zhu, H.; Zhou, Z. Screening of drought resistance indices and evaluation of drought resistance in cotton (*Gossypium hirsutum* L.). *J. Integr. Agric.* **2020**, *19*, 495–508. [[CrossRef](#)]
78. Lv, C.; Huang, Y.; Sun, W.; Yu, L.; Zhu, J. Response of rice yield and yield components to elevated [CO<sub>2</sub>]: A synthesis of updated data from FACE experiments. *Eur. J. Agron.* **2020**, *112*, 125961. [[CrossRef](#)]
79. Emamgholizadeh, S.; Parsaeian, M.; Baradaran, M. Seed yield prediction of sesame using artificial neural network. *Eur. J. Agron.* **2015**, *68*, 89–96. [[CrossRef](#)]
80. Lee, S.; Jeong, Y.; Son, S.; Lee, B. A Self-Predictable Crop Yield Platform (SCYP) Based on Crop Diseases Using Deep Learning. *Sustainability* **2019**, *11*, 3637. [[CrossRef](#)]
81. Ajay, B.C.; Bera, S.K.; Singh, A.L.; Kumar, N.; Gangadhar, K.; Kona, P. Evaluation of Genotype × Environment Interaction and Yield Stability Analysis in Peanut Under Phosphorus Stress Condition Using Stability Parameters of AMMI Model. *Agric. Res.* **2020**, 1–10. [[CrossRef](#)]
82. Finlay, K.; Wilkinson, G. The analysis of adaptation in a plant-breeding programme. *Aust. J. Agric. Res.* **1963**, *14*, 742. [[CrossRef](#)]
83. Eberhart, S.A.; Russell, W.A. Stability Parameters for Comparing Varieties 1. *Crop Sci.* **1966**, *6*, 36–40. [[CrossRef](#)]
84. Wricke, G. Über eine Methode zur Erfassung der ökologischen Streubreite in Feldversuchen. *Z. Pflanzenzüchtg* **1962**, *47*, 92–96.
85. Shukla, G.K. Some statistical aspects of partitioning genotype-environmental components of variability. *Heredity Edinb.* **1972**, *29*, 237–245. [[CrossRef](#)] [[PubMed](#)]
86. Francis, C.A.; Prager, M.; Laing, D.R.; Flor, C.A. Genotype × Environment Interactions in Bush Bean Cultivars in Monoculture and Associated with Maize. *Crop Sci.* **1978**, *18*, 237–242. [[CrossRef](#)]
87. Lin, C.S.; Binns, M.R. A superiority measure of cultivar performance for cultivar × location data. *Can. J. Plant Sci.* **1988**, *68*, 193–198. [[CrossRef](#)]
88. Karimizadeh, R.; Mohammadi, M.; Sabaghni, N.; Mahmoodi, A.A.; Roustami, B.; Seyyedi, F.; Akbari, F. GGE Biplot Analysis of Yield Stability in Multi-environment Trials of Lentil Genotypes under Rainfed Condition. *Not. Sci. Biol.* **2013**, *5*, 256–262. [[CrossRef](#)]
89. Singh, R.P.; Das, S.K.; Bhaskarrao, U.M.; Reddy, M.N. *Sustainability Index Under Different Management*; Annual Report; CRIDA: Hyderabad, India, 1990.
90. Han, X.; Hu, C.; Chen, Y.; Qiao, Y.; Liu, D.; Fan, J.; Li, S.; Zhang, Z. Crop yield stability and sustainability in a rice-wheat cropping system based on 34-year field experiment. *Eur. J. Agron.* **2020**, *113*, 125965. [[CrossRef](#)]



91. Flores, F.; Moreno, M.; Cubero, J. A comparison of univariate and multivariate methods to analyze G×E interaction. *Field Crop. Res.* **1998**, *56*, 271–286. [[CrossRef](#)]
92. Hühn, M. Beiträge zur Erfassung der phänotypischen Stabilität. I. Vorschlag einiger auf Ranginformationen beruhender Stabilitätsparameter. *EDV Medizin Biol.* **1979**, *10*, 112–117.
93. Nassar, R.; Hühn, M. Studies on Estimation of Phenotypic Stability: Tests of Significance for Nonparametric Measures of Phenotypic Stability. *Biometrics* **1987**, *43*, 45. [[CrossRef](#)]
94. Kang, M.S. A rank-sum method for selecting high-yielding, stable corn genotypes. *Cereal Res. Commun.* **1988**, *16*, 113–115.
95. Ketata, H.; Yan, S.K.; Nachit, M. Relative consistency performance across environments. In Proceedings of the International Symposium on Physiology and Breeding of Winter Cereals for Stressed Mediterranean Environments, Montpellier, France, 3–6 July 1989.
96. Fox, P.N.; Skovmand, B.; Thompson, B.K.; Braun, H.-J.; Cormier, R. Yield and adaptation of hexaploid spring triticale. *Euphytica* **1990**, *47*, 57–64. [[CrossRef](#)]
97. Li, N.; Lin, H.; Wang, T.; Li, Y.; Liu, Y.; Chen, X.; Hu, X. Impact of climate change on cotton growth and yields in Xinjiang, China. *Field Crop. Res.* **2020**, *247*, 107590. [[CrossRef](#)]
98. Harfouche, A.L.; Jacobson, D.A.; Kainer, D.; Romero, J.C.; Harfouche, A.H.; Scarascia Mugnozza, G.; Moshelion, M.; Tuskan, G.A.; Keurentjes, J.J.B.; Altman, A. Accelerating Climate Resilient Plant Breeding by Applying Next-Generation Artificial Intelligence. *Trends Biotechnol.* **2019**, *37*, 1217–1235. [[CrossRef](#)]
99. Fischer, R.; Maurer, R. Drought resistance in spring wheat cultivars. I. Grain yield responses. *Aust. J. Agric. Res.* **1978**, *29*, 897. [[CrossRef](#)]
100. Fischer, R.; Wood, J. Drought resistance in spring wheat cultivars. III.\* Yield associations with morpho-physiological traits. *Aust. J. Agric. Res.* **1979**, *30*, 1001. [[CrossRef](#)]
101. Rosielle, A.A.; Hamblin, J. Theoretical Aspects of Selection for Yield in Stress and Non-Stress Environment. *Crop Sci.* **1981**, *21*, 943–946. [[CrossRef](#)]
102. Bouslama, M.; Schapaugh, W.T. Stress Tolerance in Soybeans. I. Evaluation of Three Screening Techniques for Heat and Drought Tolerance. *Crop Sci.* **1984**, *24*, 933–937. [[CrossRef](#)]
103. Fernandez, G.C.J. Effective selection criteria for assessing stress tolerance. In Proceedings of the International Symposium on Adaptation of Vegetables and Other Food Crops in Temperature and Water Stress, Tainan, Taiwan, 13–16 August 1992; Kuo, C.G., Ed.; AVRDC Publication: Tainan, Taiwan, 1992; pp. 257–270.
104. Gavuzzi, P.; Rizza, F.; Palumbo, M.; Campanile, R.G.; Ricciardi, G.L.; Borghi, B. Evaluation of field and laboratory predictors of drought and heat tolerance in winter cereals. *Can. J. Plant Sci.* **1997**, *77*, 523–531. [[CrossRef](#)]
105. Schneider, K.A.; Rosales-Serna, R.; Ibarra-Perez, F.; Cazares-Enriquez, B.; Acosta-Gallegos, J.A.; Ramirez-Vallejo, P.; Wassimi, N.; Kelly, J.D. Improving Common Bean Performance under Drought Stress. *Crop Sci.* **1997**, *37*, 43–50. [[CrossRef](#)]
106. Farshadfar, E.; Sutka, J. Screening drought tolerance criteria in maize. *Acta Agron. Hungarica* **2002**, *50*, 411–416. [[CrossRef](#)]
107. Niazi, M.; Sadat-Noori, S.A.; Tohidfar, M.; Galuszka, P.; Mortazavian, S.M.M. Agrobacterium-mediated genetic transformation of ajowan (*Trachyspermum ammi* (L.) Sprague): An important industrial medicinal plant. *Ind. Crops Prod.* **2019**, *132*, 29–40. [[CrossRef](#)]
108. Niazi, M.; Shariatpanahi, M.E. In vitro-based doubled haploid production: Recent improvements. *Euphytica* **2020**, *216*, 69. [[CrossRef](#)]
109. Marchetti, C.F.; Ugena, L.; Humplík, J.F.; Polák, M.; Čavar Zeljković, S.; Podlešáková, K.; Fürst, T.; De Diego, N.; Spíchal, L. A Novel Image-Based Screening Method to Study Water-Deficit Response and Recovery of Barley Populations Using Canopy Dynamics Phenotyping and Simple Metabolite Profiling. *Front. Plant Sci.* **2019**, *10*. [[CrossRef](#)]
110. Barbedo, J.G.A. Detection of nutrition deficiencies in plants using proximal images and machine learning: A review. *Comput. Electron. Agric.* **2019**, *162*, 482–492. [[CrossRef](#)]
111. 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](#)]
112. Picon, A.; Alvarez-Gila, A.; Seitz, M.; Ortiz-Barredo, A.; Echazarra, J.; Johannes, A. Deep convolutional neural networks for mobile capture device-based crop disease classification in the wild. *Comput. Electron. Agric.* **2019**, *161*, 280–290. [[CrossRef](#)]

113. Agarwal, M.; Sinha, A.; Gupta, S.K.; Mishra, D.; Mishra, R. Potato crop disease classification using convolutional neural network. In *Smart Systems and IoT: Innovations in Computing*; Somani, A.K., Shekhawat, R.S., Mundra, A., Srivastava, S., Verma, V.K., Eds.; Smart Innovation, Systems and Technologies; Springer: Singapore, 2020; Volume 141, ISBN 978-981-13-8405-9.
114. Anagnostis, A.; Asiminari, G.; Papageorgiou, E.; Bochtis, D. A Convolutional Neural Networks Based Method for Anthracnose Infected Walnut Tree Leaves Identification. *Appl. Sci.* **2020**, *10*, 469. [\[CrossRef\]](#)
115. Khamparia, A.; Saini, G.; Gupta, D.; Khanna, A.; Tiwari, S.; de Albuquerque, V.H.C. Seasonal Crops Disease Prediction and Classification Using Deep Convolutional Encoder Network. *Circuits Syst. Signal Process.* **2020**, *39*, 818–836. [\[CrossRef\]](#)
116. Sibiya, M.; Sumbwanyambe, M. A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks. *AgriEngineering* **2019**, *1*, 9. [\[CrossRef\]](#)
117. Kearsey, M.; Pooni, H. *The Genetical Analysis of Quantitative Traits*; Stanley Thornes Ltd: Cheltenham, UK, 1998.
118. Griffing, B. Concept of General and Specific Combining Ability in Relation to Diallel Crossing Systems. *Aust. J. Biol. Sci.* **1956**, *9*, 463. [\[CrossRef\]](#)
119. Hayman, B.I. The Theory and Analysis of Diallel Crosses. II. *Genetics* **1958**, *43*, 63–85. [\[PubMed\]](#)
120. Jinks, J.L. The Analysis of Continuous Variation in a Diallel Cross of *Nicotiana Rustica* Varieties. *Genetics* **1954**, *39*, 767–788. [\[PubMed\]](#)
121. Kempthorne, O. *An Introduction to Genetic Statistics*; John Wiley & Sons Inc.: New York, NY, USA, 1957.
122. Comstock, R.E.; Robinson, H.F. The Components of Genetic Variance in Populations of Biparental Progenies and Their Use in Estimating the Average Degree of Dominance. *Biometrics* **1948**, *4*, 254. [\[CrossRef\]](#) [\[PubMed\]](#)
123. Opsahl, B. The Discrimination of Interactions and Linkage in Continuous Variation. *Biometrics* **1956**, *12*, 415. [\[CrossRef\]](#)
124. Kearsey, M.J.; Jinks, J.L. A general method of detecting additive, dominance and epistatic variation for metrical traits I. Theory. *Heredity Edinb.* **1968**, *23*, 403–409. [\[CrossRef\]](#)
125. Dezfouli, P.M.; Sedghi, M.; Shariatpanahi, M.E.; Niazian, M.; Alizadeh, B. Assessment of general and specific combining abilities in doubled haploid lines of rapeseed (*Brassica napus* L.). *Ind. Crop. Prod.* **2019**, *141*, 111754. [\[CrossRef\]](#)
126. Khaki, S.; Khalilzadeh, Z.; Wang, L. Predicting yield performance of parents in plant breeding: A neural collaborative filtering approach. *PLoS ONE* **2020**, *15*, e0233382. [\[CrossRef\]](#)
127. Niazian, M. Application of genetics and biotechnology for improving medicinal plants. *Planta* **2019**, *249*, 953–973. [\[CrossRef\]](#)
128. Ayuso, M.; García-Pérez, P.; Ramil-Rego, P.; Gallego, P.P.; Barreal, M.E. In vitro culture of the endangered plant *Eryngium viviparum* as dual strategy for its ex situ conservation and source of bioactive compounds. *Plant Cell Tissue Organ Cult.* **2019**, *138*, 427–435. [\[CrossRef\]](#)
129. Sugimoto, K.; Temman, H.; Kadokura, S.; Matsunaga, S. To regenerate or not to regenerate: Factors that drive plant regeneration. *Curr. Opin. Plant Biol.* **2019**, *47*, 138–150. [\[CrossRef\]](#) [\[PubMed\]](#)
130. Prasad, V.S.S.; Gupta, S.D. Applications and potentials of artificial neural networks in plant tissue culture. In *Plant Tissue Culture Engineering*; Gupta, S.D., Ibaraki, Y., Eds.; Focus on Biotechnology; Springer: Dordrecht, The Netherlands, 2008; Volume 6, ISBN 978-1-4020-3594-4.
131. Niazian, M.; Noori, S.A.S.; Galuszka, P.; Tohidfar, M.; Mortazavian, S.M.M. Genetic stability of regenerated plants via indirect somatic embryogenesis and indirect shoot regeneration of *Carum copticum* L. *Ind. Crops Prod.* **2017**, *97*, 330–337. [\[CrossRef\]](#)
132. Orłowska, R.; Bednarek, P.T. Precise evaluation of tissue culture-induced variation during optimisation of in vitro regeneration regime in barley. *Plant Mol. Biol.* **2020**, *103*, 33–50. [\[CrossRef\]](#) [\[PubMed\]](#)
133. Phillips, G.C.; Garda, M. Plant tissue culture media and practices: An overview. *Vitr. Cell. Dev. Biol. Plant* **2019**, *55*, 242–257. [\[CrossRef\]](#)
134. Alanagh, E.N.; Garoosi, G.; Haddad, R.; Maleki, S.; Landín, M.; Gallego, P.P. Design of tissue culture media for efficient *Prunus* rootstock micropropagation using artificial intelligence models. *Plant Cell Tissue Organ Cult.* **2014**, *117*, 349–359. [\[CrossRef\]](#)
135. Hesami, M.; Naderi, R.; Tohidfar, M. Modeling and Optimizing Medium Composition for Shoot Regeneration of *Chrysanthemum* via Radial Basis Function-Non-dominated Sorting Genetic Algorithm-II (RBF-NSGAI). *Sci. Rep.* **2019**, *9*, 18237. [\[CrossRef\]](#)

136. Sadat Noori, S.A.; Norouzi, M.; Karimzadeh, G.; Shirkool, K.; Niazian, M. Effect of colchicine-induced polyploidy on morphological characteristics and essential oil composition of ajowan (*Trachyspermum ammi* L.). *Plant Cell Tissue Organ Cult.* **2017**, *130*, 543–551. [\[CrossRef\]](#)
137. Castillo, A.M.; Cistué, L.; Vallés, M.P.; Soriano, M. Chromosome Doubling in Monocots. In *Advances in Haploid Production in Higher Plants*; Springer: Dordrecht, The Netherlands, 2009; pp. 329–338.
138. Niazian, M.; Sadat Noori, S.A.; Galuszka, P.; Mortazavian, S.M.M. Tissue culture-based Agrobacterium-mediated and in planta transformation methods. *Czech J. Genet. Plant Breed.* **2017**, *53*, 133–143. [\[CrossRef\]](#)
139. Matias, F.I.; Caraza-Harter, M.V.; Endelman, J.B. FIELDimageR: An R package to analyze orthomosaic images from agricultural field trials. *Plant Phenome J.* **2020**, *3*, e20005. [\[CrossRef\]](#)
140. Tsaftaris, S.A.; Minervini, M.; Scharf, H. Machine Learning for Plant Phenotyping Needs Image Processing. *Trends Plant Sci.* **2016**, *21*, 989–991. [\[CrossRef\]](#)
141. Ubbens, J.R.; Stavness, I. Deep Plant Phenomics: A Deep Learning Platform for Complex Plant Phenotyping Tasks. *Front. Plant Sci.* **2017**, *8*, 1190. [\[CrossRef\]](#) [\[PubMed\]](#)
142. Li, X.; He, Y. Discriminating varieties of tea plant based on Vis/NIR spectral characteristics and using artificial neural networks. *Biosyst. Eng.* **2008**, *99*, 313–321. [\[CrossRef\]](#)
143. Khoshroo, A.; Arefi, A.; Masoumiasl, A.; Jowkar, G.H. Classification of wheat cultivars using image processing and artificial neural networks. *Agric. Commun.* **2014**, *2*, 17–22.
144. Mahlein, A.-K. Plant Disease Detection by Imaging Sensors—Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Dis.* **2016**, *100*, 241–251. [\[CrossRef\]](#) [\[PubMed\]](#)
145. Maes, W.H.; Steppe, K. Perspectives for Remote Sensing with Unmanned Aerial Vehicles in Precision Agriculture. *Trends Plant Sci.* **2019**, *24*, 152–164. [\[CrossRef\]](#) [\[PubMed\]](#)
146. Khadka, K.; Earl, H.J.; Raizada, M.N.; Navabi, A. A Physio-Morphological Trait-Based Approach for Breeding Drought Tolerant Wheat. *Front. Plant Sci.* **2020**, *11*, 715. [\[CrossRef\]](#)
147. Singh, A.; Jones, S.; Ganapathysubramanian, B.; Sarkar, S.; Mueller, D.; Sandhu, K.; Nagasubramanian, K. Challenges and Opportunities in Machine-Augmented Plant Stress Phenotyping. *Trends Plant Sci.* **2020**. [\[CrossRef\]](#)
148. Pan-pan, H.; Wei-xu, L.; Hui-hui, L.; Zeng-xu, X. In vitro induction and identification of autotetraploid of *Bletilla striata* (Thunb.) Reichb.f. by colchicine treatment. *Plant Cell Tissue Organ Cult.* **2018**, *132*, 425–432. [\[CrossRef\]](#)
149. Ochatt, S.J.; Patat-Ochatt, E.M.; Moessner, A. Ploidy level determination within the context of in vitro breeding. *Plant Cell Tissue Organ Cult.* **2011**, *104*, 329–341. [\[CrossRef\]](#)
150. Ahmadi, B.; Ebrahimzadeh, H. In vitro androgenesis: Spontaneous vs. artificial genome doubling and characterization of regenerants. *Plant Cell Rep.* **2020**, *39*, 299–316. [\[CrossRef\]](#)
151. Santeramo, D.; Howell, J.; Ji, Y.; Yu, W.; Liu, W.; Kelliher, T. DNA content equivalence in haploid and diploid maize leaves. *Planta* **2020**, *251*, 30. [\[CrossRef\]](#) [\[PubMed\]](#)
152. Blonder, B.; Graae, B.J.; Greer, B.; Haagsma, M.; Helsen, K.; Kapás, R.E.; Pai, H.; Rieksta, J.; Sapena, D.; Still, C.J.; et al. Remote sensing of ploidy level in quaking aspen (*Populus tremuloides* Michx.). *J. Ecol.* **2020**, *108*, 175–188. [\[CrossRef\]](#)

