Machine Learning and Vegetable Science

Along with essential nutrients and trace elements, vegetables provide raw materials for the food processing industry. Despite this, plant diseases and unfavorable weather patterns continue to threaten the delicate balance between vegetable production and consumption. It is critical to utilize machine learning (ML) in this setting because it provides context for decision-making related to breeding goals.

1. ML Models

ML approaches are powerful tools that can solve complex nonlinear problems on their own using sensor data and this allows for more informed decision-making and actions to be taken in practical scenarios with nominal human intervention. ML techniques are constantly evolving and are used in almost every domain. Their applications, however, have some fundamental restrictions. Data quality, model representation, and input-target variable correlations all have an impact on prediction accuracy. Models used for machine learning tasks are divided into three main types, as shown in **Figure 1**.
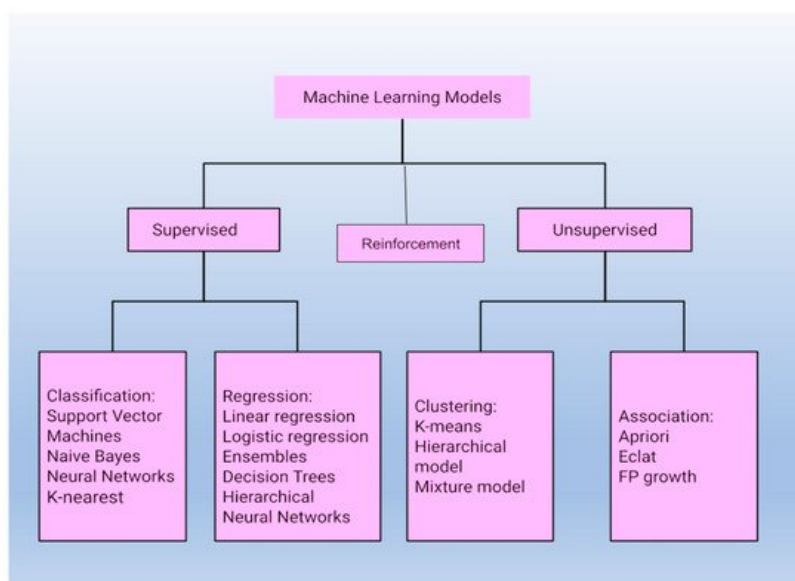

Figure 1. Classification of models used for machine learning tasks into three major categories.

Regression algorithms, a class of supervised ML technique, generally constitute linear and logistic regression models. Linear regression models primarily represent a linear correlation between independent and dependent variables [1], producing a straight line graph. Logistic regression, on the other hand, produces a non-linear curve with output values lying in between 0 and 1. Furthermore, complex regression models have also been created, such as ordinary least squares, multiple linear regression, locally linearized splines, and cubist regression [2].

Another ML approach, artificial neural network (ANN), is a type of information processing system that works in a similar way to biological neural networks. This method is used to recognize non-linear and complex functions. In fact, the use of ANN supervised learning techniques can help with both regression and classification problems [3]. Further, deep ANNs, also known as deep neural networks (DNNs) or deep learning (DL), allow computational models with multiple processing layers to learn complex data [4]. A popular DL model is the convolutional neural network (CNN), which extracts feature maps by carrying out convolutions in the image domain.

Support vector machines (SVMs) are also regularly used for vegetable crops [5]. SVM is essentially a binary classifier that creates a linear separating hyperplane to classify data instances. SVM regression algorithms are generally used to predict a continuous response, finding a model that deviates from the calculated data by small amounts; rather than using a hyperplane to separate data, it uses parametric models to detect minor differences [6]. Data with a huge number of predictors fit good with it. Yield and sensor data forecasting are two possible supervised learning applications in agriculture [7].

Bayesian models (BM), a type of probabilistic graphical models in which Bayesian inference is used to initiate the analysis, also constitute a major class of supervised ML models [8]. Bayesian inference, unlike most ML algorithms, requires only a small number of training samples [9]. The Bayes' Theorem, which serves as the foundation for BM, is represented by the equation: P(A|B) = P(B|A) [P(A)/P(B)] [8]. This equation is used to compute the posterior probability (P(A|B)) based on the prior probability (P(A)) and the information gathered from the data. P(B|A) denotes the likelihood of the observation B.

Further, Decision trees (DT) have also found their applications in data analysis for vegetable crops. DTs are used to organize the dataset into smaller homogeneous subsets (sub-populations) while simultaneously creating a linked tree graph [10]. However, the model can be modified to make it simpler by eliminating branches. DT is ideal for applications that don't require high predictive accuracy. When compared to other ML methods, simple regression trees do not perform satisfactorily. However, out of the different tree-based approaches, random forests (RFs) have been recognized as the most effective and widespread ML approach [11].
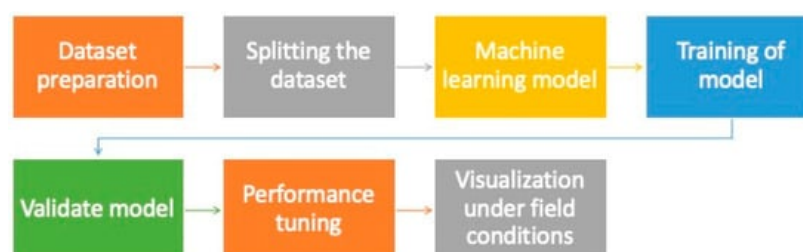
Another ML algorithm that needs to be mentioned is ensemble learning (EL). By constructing a linear aggregate of a "base learning algorithm," EL models aim to improve any model fitting technique's predictive performance [12]. Further, a common method such as the RF (ensemble or grouping of DTs) algorithm avoids overfitting by lowering the variance in DT [13].

2. Tasks Employing ML in Vegetables

2.1. Assessment of Seed Quality

Seed quality is a crucial factor in vegetable production because it affects the yield directly [14][15][16]. For example, calculating the germination percentage often necessitates professional technicians manually counting and grading germinating seedlings [17][18]. Further, seed sowing quality is determined by seed composition, kernel maturity, insect infestation, diseases, cleanliness, and germination ability, and is linked to post-sowing germination and growth conditions. Plant genetic purity can be determined using molecular recognition, DNA analysis, isotope fingerprinting, and mineral element analysis [19][20]. In addition, in order to determine the seed vigour and germination, several techniques such as high-performance liquid chromatography, tetrazolium tests, and conductivity tests are used [21][22]. Although the majority of these chemical and physical techniques have high precision and reliability, they come at a high cost, take a long time, and require a lot of operators.

Plant breeding that uses high-quality seeds reduces the cost of field experiments while increasing the chances of finding a better crop variety. However, these procedures are limited by time, subjectivity, and the destructive nature of seed quality assessment. The present scenario demands quick, reliable, non-destructive, and objective methods for detecting seed quality. It has been observed that variations in the chemical composition and internal anatomical features of seeds are often linked to the loss of viability and vigor, however, these changes are difficult to detect when examined visually [22][23]. Meanwhile, data on complex seed quality traits have been successfully collected using spectrometry and X-ray imaging techniques [24]. The agricultural industry has been transformed significantly by recent advances in ML algorithms, which serve as the foundation for developing models to classify products, particularly seed quality attributes. The application of such ML models for assessment of seed quality before germination can significantly help in increasing vegetable production with desired traits. The development of a machine learning model for assessing seed quality has various steps, as shown in **Figure 2**.

Figure 2. Step in establishment of a machine learning model for seed quality assessment.

2.2. Disease Detection and Control

The most common pest and disease control method adopted in vegetable production is to spray pesticides evenly across the cropping area [25][26][27]. While this method is relatively effective, it comes at a significant financial and environmental cost. Residues in crop products, groundwater contamination, food chain contamination, and effects on local wildlife and ecosystems are just a few examples of environmental consequences. Plant diseases have long been a major concern in vegetable production due to their ability to reduce crop quality and, subsequently, the production [28][29][30]. They can cause severe damage to entire areas of planted crops, resulting in significant financial loss and a considerable impact on the agricultural economy, especially in developing countries where a single crop or a few crops are the primary sources of income [31][32].

In order to avoid major losses, various methods for diagnosing disease have been developed. The precise identification of causative agents is now possible thanks to advances in molecular biology and immunology. Many farmers, however, are unable to implement these methods due to the requirement of extensive domain knowledge or a significant amount of money and resources. However, since these farmers bear the responsibility of feeding a large percentage of the world's population, extensive research has been carried out in order to develop methods that are both accurate and accessible to the vast majority of farmers.

Precision agriculture uses cutting-edge technology to help farmers make better decisions for detection and control of diseases that employ ML to target agrochemical inputs [33][34]. With the progression of modern digital technologies, a large amount of data is collected in real time, and various ML algorithms are used to make optimal decisions [35][36]. Further, vegetable production has been impacted by a recent surge in DL methods. Novel solutions may emerge as a result of advances in computer vision and artificial intelligence which are far more effective and accurate than traditional methods at making predictions, thereby allowing for better decision-making. DL methods are now used to solve complex problems related to plant diseases in a reasonable amount of time owing to advancements in hardware technology [37][38]. Examples of ML applications in the diagnosis, prevention, and control of disease in vegetable crops are provided in **Table 1**. However, there is still room for improvement in this area, particularly in decision-support systems that aid in the conversion of large amounts of data into actionable recommendations.

Table 1. Examples of application of ML in disease detection, prevention and control [39].

Application	ML Tool
Estimation of *Phytophthora infestans* infection in tomato under field condition.	Neural Network
Foliar diseases of sugar beet in glasshouse conditions.	Support Vector Machine
Detection of *Oidium neolycopersici* infestation in tomato.	Support Vector Machine
Bacterial infection in *Cucumis melo* under glasshouse conditions.	Logistic Regression, Support Vector Machine, Neural Network
Disease detection in plant species including vegetables.	Convolutional Neural Network
Gene regulatory network of the pathogenic fungus *Fusarium graminearum* constructed from hundreds of transcriptomic datasets.	Bayesian network inference
EffectiveT3: Identification of N-terminal signal peptide.	Naïve Bayes
DeepT3: Identification of bacterial type III secreted effectors.	Deep Convolutional Neural Network
T4SEpre: prediction of bacterial type IV secreted proteins.	Support Vector Machine
Bastion6: prediction of bacterial type VI secreted proteins.	Support Vector Machine
EffectorP: fungal effector prediction.	Naïve Bayes, Ensemble Learner
ApoplastP: localization of the effector proteins.	Random Forest
LOCALIZER: localization of plant proteins	Support Vector Machine

2.3. Prediction of Climatic Variations

The environment (climate), agricultural operations in vegetable production (sowing, cultivation, and harvesting), and plant genotype all influence crop yield and productivity [40][41][42]. The interactions and relationships (direct and/or indirect) among these factors create a complex situation in which potential plant yield is determined. Year-to-year variations in a genotype's yield and phenotypic trait are caused by environmental variations and genotype × environment interaction (GEI). Stability analysis, which estimates genotypes' relative performance across different environments, is a perfect solution to these yearly variations [43]. The use of deterministic, biophysical crop models for yield modelling for the purpose of assessing the impact of climate change accounts for a significant portion of the work in this area [44][45]. On the other hand, statistical models outperform them when it comes to predicting over larger spatial scales. Statistical models have been used in a large body of literature to demonstrate a strong link between extreme heat and poor crop performance because their objectives are primarily focused on outcome prediction rather than inference into the nature of the mechanistic processes generating those outcomes. In an ANN model, plant growth indices could be used as dependent variables while climate variables could be used as independent variables [46][47].

The linear and nonlinear relationships between variables can then be considered using powerful ANN models. Deep phenotyping combined with AI is a useful tool for figuring out how plants interact with their surroundings [48][49][50]. Further, semiparametric neural networks (SNN) are a novel way to combine DNNs with parametric statistical models [51]. When used as a crop yield modelling framework, the SNN outperforms everything else in terms of out-of-sample predictive performance [52][53]. This, when combined with a number of complementary methods, outperforms both existing parametric approaches and fully nonparametric neural networks in terms of efficiency and, ultimately, performance [54].

2.4. Crop Monitoring and Yield Prediction

Vegetable production benefits greatly from ML technology, as it makes it easier to monitor, scan, and analyze crops by providing high-quality images [55][56]. This is highly useful for assessing crop health and determining crop progress. Farmers, for example, can use the images provided by this technology to determine whether or not their crops are ready to harvest. Farmers can use DL and other ML techniques to assess the state of their soil [57][58]. DL is also used to determine the best times for planting and harvesting and how water and nutrients must be managed [59]. This, of course, enhances farming efficiency, and the return on investment (ROI) from specific crops can be predicted by considering their price and market margin [60][61]. With high-performance computers becoming more common, ML techniques becoming more popular, and satellite imagery data becoming more widely available, there is a chance to develop fast, accurate, and reliable methods for generating crop yield maps [62][63].

Since there are various crop growth-related biochemical and biophysical characteristics that must be measured at a fine scale to assist in irrigation, fertilizer, and pesticide application decisions, such as leaf nitrogen concentration (N), leaf area index (LAI), and above-ground biomass (AGB), we must track these aspects carefully [64]. Furthermore, high-throughput plant phenotyping creates an urgent need for precision crop monitoring that is both cost-effective and non-destructive [65][66]. Crop monitoring in vegetable production has traditionally relied on field-based surveying and sampling, as well as laboratory-based analyses; however, these methods are time-consuming, can be destructive, and are not practical for large-scale applications. Alternatively, the biochemical and biophysical traits have been estimated using satellite remote sensing [67][68]. However, satellite remote sensing applications at fine spatial and temporal scales are limited by insufficient spatial resolution and revisiting frequencies [69]. Also, the atmospheric conditions and soil background effects may limit the optical satellite data. Furthermore, the lack of three-dimensional (3D) canopy structural information and asymptotic saturation phenomena inherent in optical spectral data limit its use for crop monitoring, especially in dense and diverse canopies at advanced stages of development [68][69][70]. Recent advances in ML and, in particular, deep learning (DL), have enabled the development of several new analytical tools. In this direction, a recent study predicted the yield of tomatoes employing ML and deep learning techniques under controlled greenhouses and in uncontrolled greenhouses. The authors used recurrent neural networks (RNN) for prediction formulations based on the Long Short-Term Memory (LSTM) neuron model of the deep learning approach. The RNN architecture calculates the parameters in conjunction with previous yield, growth, and stem diameter values and microclimate conditions [71][72].

2.5. ML and Vegetable Breeding

Understanding a complex trait, such as yield, as a function of genetic, phenotypic, and environmental data is one of the most critical targets of plant science and breeding. ML and other approaches are being used to classify quantitative trait loci (QTLs) [73] or genomic regions related to phenotypes. In this sense, genetic associations between traits, which calculate the degree of overlap between genetic signals, are concerned. As well, they can estimate trait values for new genotypes with only marker data genomic prediction (GP). Vegetable breeders regularly use genomic selection

approaches, which entails selecting material based on GPs rather than phenotypic values and marker-assisted selection, which necessitates QTL mapping [74]. Breeders are particularly interested in the genes that underpin QTLs. A variety of architectures for GP, with the help of DL approaches, have been produced [75]. Although some researchers have investigated ML methods for QTL mapping (primarily for pre-screening), their use is constrained because practitioners often need *p* values or other confidence measures to validate the outcomes.

Random-effects are used to estimate GP and effect sizes simultaneously [76]. In contrast to parametric random-effects models, ML has several advantages. First, manually designed features are needed where secondary characteristics, such as picture meaning, are more complicated. Second, ML methods (especially DL) could be more resilient in situations where common assumptions such as Gaussianity are breached, such as for traits calculated on an ordinal scale. Third, when a portion of the genetic variation is non-additive, ML may increase accuracy [77]. However, ML falls short of random-effects models for certain characteristics. In general, plant breeding in the private sector still relies heavily on rigorous yield assessment in several environments. With a few variations, phenotypes may be very multidimensional, and there are typically hundreds of genotypes. Any recent developments in ML could help solve this problem. It is a big challenge to predict genotype rating within each new setting based on environmental variables that characterize G × E interactions; random-effects approaches were used to show that this is possible using ML approaches [78]. More analysis is required to define the most significant environmental variables and realistic data-driven environmental features. The question of when and how they will evolve GP for the target trait, now that these traits are available, emerges. This is true for a single secondary trait with a high enough heritability and genetic resemblance to the target trait. Several authors have used omics, environmental, or management data to forecast yield in the absence of marker data; see, for example, [79]. While such models cannot make genomic choices, they can help policymakers and farmers make more educated decisions. These algorithms propose causal relationships between phenotypes that are more closely related to the results, and they may also incorporate previously defined functional relationships. These approaches' importance lies in their ability to forecast treatment conditions, such as what will happen if different situations or coping mechanisms are employed, or if a gene is silenced.

2.6. ML and Vegetable Biotechnology

Agrobacterium-mediated gene synthesis is a well-known technique for plant gene transformation and genetic alteration [80]. For effective gene transfer, the Agrobacterium strain, the period of inoculation, and select antibiotic concentrations must be fine-tuned [81]. Moreover, it was determined that resistance to Agrobacterium-assisted gene transformation is easily observable using ML algorithms [82]. Polyploidy is often utilized to increase the productivity and vigor of plants [83]. This results in a close correlation between the plant genotype and the antimitotic agent in artificial polyploidy induction [84].

Many in vitro-based breeding strategies depend on in vitro regeneration, which has a broad variety of applications in plant breeding [83]. Micropropagation (proliferation) and in vitro regeneration have clear effects on both in-situ and ex-situ conservation. In vitro culture is an effective tool for widespread reproduction, germplasm survival, and bioactive compound processing in several endangered uncommon plant species, including medicinal plants [85]. A variety of variables influence the fate of cultured cells' in vitro plant regeneration [86]. The mixture and interactions between these variables are responsible for the in vitro plant regeneration process's multifactorial nature. When other factors are introduced, the scenario becomes incredibly difficult to comprehend. Manipulation of the basal medium has been used as a promotion strategy to enhance in vitro studies' efficiency [87]. The cytokinin/auxin ratio is also essential in in vitro tests. The ANN model correctly estimated the number and length of microshoots. In the ANN's sensitivity study, the immersion time was deemed to be the most critical element affecting pollution level and explant viability. Further, ANNs are often used to forecast plantlet growth from embryos, calculation of cell culture biomass, simulation of temperature distribution in a culture vessel, identification and calculation of in vitro induced shoot weight, and clustering of in vitro regenerated plantlets [88].

2.7. ML and Vegetable Genomics

Understanding the movement of information is vital for the study of vegetable sciences and crop improvement, but how to do so is a mystery. This chasm is being filled by advances in two fields of analysis in particular. Association research between molecular phenotype and terminal phenotype benefits from a shorter recognition rail and includes less information transfer than genome-wide connection studies and transcriptome wide-interaction studies. Interaction imaging was used to identify genetic loci associated with molecular or terminal characteristics in natural plant populations [89]. Due to widespread linkage imbalance among the neighboring variants and impeding genetic enhancement of plants by genome editing, variants underlying phenotypic variance are difficult to distinguish. On the other hand, advancements in molecular biology over the last half-century have assisted in the discovery of many molecular pathways that control the

flow of data from DNA to RNA and protein, and the assortment of such data is determined by a number of omics methods based on sophisticated sequencing techniques [90].

CNNs have at least one convolutional layer, which allows them to derive features from a continuous signal (for example, weather data as a time series, a plant image, or a DNA/RNA sequence). A DNA/RNA sequence of N base pairs can be used to train a CNN, and it can be defined as a one-hot encoded 4 N matrix [91]. Even if local motifs exist in separate parts of the input, CNNs can catch them. Convolutional layers often reduce the number of weights that must be learned as compared to fully connected layers. Many CNN uses in plant biology are offered as an interactive tutorial for building a CNN to investigate DNA-binding motifs [50]. As a consequence of this operation, recurrent neural networks (RNNs) acquire memory capabilities. When dealing with time series inputs, RNNs may also handle a range of input sizes, which is beneficial. When using ML to solve problems in genomics [92], there are a few critical aspects to bear in mind. The model should generalize well, which means it should act consistently between test and training sets. Several variables, such as model complexity, high dimensionality, and so on, can lead to overfitting. Large-scale phenotyping trials are often expensive; estimating genomic variant phenotypes almost always costs more than the amount of plant genotypes [93]. Overfitting is often concealed and ignored when confronted with genomics problems.

As a result, it seems plausible to anticipate that selecting causative variations can be performed by combining models that "understand" the information flow from DNA to molecular phenotypes with interaction mapping investigations that connect molecular phenotypes to behavioral traits. Indeed, in human genetics, such a structure has been shown not only to be probable but also successful in revealing variants (including rare alleles) at the root of certain genetic disorders [94]. On the other hand, the vegetable community is yet to benefit from this trend entirely. DL models have made significant progress in developing molecular phenotype prediction and we believe that such a device would help identify deleterious and adaptive variants in the genome, which would be essential for potential crop editing-based genetic enhancement.

Any protein's function is directly linked to its tertiary structure. Secondary composition, transmembrane topology, signal peptides, and enzyme dynamics are some of the protein properties that can be integrated and studied to reveal the tertiary structure. Google's Alpha Fold recently made news as it used AI to predict a protein's tertiary structure. On the other hand, DL algorithms have shown promise in a variety of fields [95]. Still, their utility in predicting protein–protein interactions (PPI) has been constrained by inadequate coverage and noisy data.

References

1. Darlington, R.B.; Hayes, A.F. Regression Analysis and Linear Models: Concepts, Applications, and Implementation; Guilford Publications: New York, NY, USA, 2016.

2. Ali, I.; Greifeneder, F.; Stamenkovic, J.; Neumann, M.; Notarnicola, C. Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data. Remote Sens. 2015, 7, 16398–16421.

3. Sathya, R.; Abraham, A. Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. Int. J. Adv. Res. Artif. Intell. 2013, 2, 34–38.

4. Koutsoukas, A.; Monaghan, K.J.; Li, X.; Huan, J. Deep-Learning: Investigating Deep Neural Networks Hyper-Parameters and Comparison of Performance to Shallow Methods for Modeling Bioactivity Data. J. Cheminform. 2017, 9, 1–13.

5. Luo, S.-T.; Cheng, B.-W.; Hsieh, C.-H. Prediction Model Building with Clustering-Launched Classification and Support Vector Machines in Credit Scoring. Expert Syst. Appl. 2009, 36, 7562–7566.

6. Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. A Comparative Assessment of Support Vector Regression, Artificial Neural Networks, and Random Forests for Predicting and Mapping Soil Organic Carbon Stocks across an Afromontane Landscape. Ecol. Indic. 2015, 52, 394–403.

7. Mekonnen, Y.; Namuduri, S.; Burton, L.; Sarwat, A.; Bhansali, S. Machine Learning Techniques in Wireless Sensor Network Based Precision Agriculture. J. Electrochem. Soc. 2019, 167, 037522.

8. Korner-Nievergelt, F.; Roth, T.; Von Felten, S.; Guélat, J.; Almasi, B.; Korner-Nievergelt, P. Bayesian Data Analysis in Ecology Using Linear Models with R, BUGS, and Stan; Academic Press: Cambridge, MA, USA, 2015.

9. Barber, D. Bayesian Reasoning and Machine Learning; Cambridge University Press: Cambridge, UK, 2012.

10. Liakos, K.G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. Sensors 2018, 18, 2674.

11. Chen, X.; Ishwaran, H. Random Forests for Genomic Data Analysis. Genomics 2012, 99, 323–329.

12. Onan, A.; Korukoğlu, S.; Bulut, H. A Hybrid Ensemble Pruning Approach Based on Consensus Clustering and Multi-Objective Evolutionary Algorithm for Sentiment Classification. Inf. Process. Manag. 2017, 53, 814–833.

13. Watts, J.D.; Lawrence, R.L. Merging Random Forest Classification with an Object-Oriented Approach for Analysis of Agricultural Lands. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. 2008, 37, 2006–2009.

14. Afzal, I.; Shabir, R.; Rauf, S. Seed production technologies of some major field crops. In Agronomic Crops; Springer: Berlin/Heidelberg, Germany, 2019; pp. 655–678.

15. Lamichhane, J.R.; Debaeke, P.; Steinberg, C.; You, M.P.; Barbetti, M.J.; Aubertot, J.-N. Abiotic and Biotic Factors Affecting Crop Seed Germination and Seedling Emergence: A Conceptual Framework. Plant Soil 2018, 432, 1–28.

16. Ratnadass, A.; Fernandes, P.; Avelino, J.; Habib, R. Plant Species Diversity for Sustainable Management of Crop Pests and Diseases in Agroecosystems: A Review. Agron. Sustain. Dev. 2012, 32, 273–303.

17. Shamshiri, R.R.; Weltzien, C.; Hameed, I.A.; Yule, I.J.; Grift, T.E.; Balasundram, S.K.; Pitonakova, L.; Ahmad, D.; Chowdhary, G. Research and Development in Agricultural Robotics: A Perspective of Digital Farming. Int. J. Agric. Bio. Eng. 2018, 11, 1–14.

18. Rahman, A.; Cho, B.-K. Assessment of Seed Quality Using Non-Destructive Measurement Techniques: A Review. Seed Sci. Res. 2016, 26, 285–305.

19. Danezis, G.P.; Tsagkaris, A.S.; Camin, F.; Brusic, V.; Georgiou, C.A. Food Authentication: Techniques, Trends & Emerging Approaches. Trac Trends Anal. Chem. 2016, 85, 123–132.

20. Wadood, S.A.; Boli, G.; Xiaowen, Z.; Hussain, I.; Yimin, W. Recent Development in the Application of Analytical Techniques for the Traceability and Authenticity of Food of Plant Origin. Microchem. J. 2020, 152, 104295.

21. Dell'Aquila, A. Perspectives in Probing Seed Germination and Vigour. Seed Sci. Biotechnol. 2008, 2, 1–14.

22. ElMasry, G.; Mandour, N.; Al-Rejaie, S.; Belin, E.; Rousseau, D. Recent Applications of Multispectral Imaging in Seed Phenotyping and Quality Monitoring—An Overview. Sensors 2019, 19, 1090.

23. Dell'Aquila, A. Development of Novel Techniques in Conditioning, Testing and Sorting Seed Physiological Quality. Seed Sci. Technol. 2009, 37, 608–624.

24. De Medeiros, A.D.; da Silva, L.J.; Ribeiro, J.P.O.; Ferreira, K.C.; Rosas, J.T.F.; Santos, A.A.; da Silva, C.B. Machine Learning for Seed Quality Classification: An Advanced Approach Using Merger Data from FT-NIR Spectroscopy and X-Ray Imaging. Sensors 2020, 20, 4319.

25. Matthews, G. Pesticide Application Methods; John Wiley & Sons: Hoboken, NJ, USA, 2008.

26. Obopile, M.; Munthali, D.C.; Matilo, B. Farmers' Knowledge, Perceptions and Management of Vegetable Pests and Diseases in Botswana. Crop Prot. 2008, 27, 1220–1224.

27. Schreinemachers, P.; Balasubramaniam, S.; Boopathi, N.M.; Ha, C.V.; Kenyon, L.; Praneetvatakul, S.; Sirijinda, A.; Le, N.T.; Srinivasan, R.; Wu, M.-H. Farmers' Perceptions and Management of Plant Viruses in Vegetables and Legumes in Tropical and Subtropical Asia. Crop Prot. 2015, 75, 115–123.

28. Bisbis, M.B.; Gruda, N.; Blanke, M. Potential Impacts of Climate Change on Vegetable Production and Product Quality–A Review. J. Clean. Prod. 2018, 170, 1602–1620.

29. Castañé, C.; Arnó, J.; Gabarra, R.; Alomar, O. Plant Damage to Vegetable Crops by Zoophytophagous Mirid Predators. Biol. Control 2011, 59, 22–29.

30. Rubatzky, V.E.; Yamaguchi, M. World Vegetables: Principles, Production, and Nutritive Values; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.

31. Strange, R.N.; Scott, P.R. Plant Disease: A Threat to Global Food Security. Annu. Rev. Phytopathol. 2005, 43, 83–116.

32. Vurro, M.; Bonciani, B.; Vannacci, G. Emerging Infectious Diseases of Crop Plants in Developing Countries: Impact on Agriculture and Socio-Economic Consequences. Food Secur. 2010, 2, 113–132.

33. Mondal, P.; Basu, M.; Bhadoria, P.B.S. Critical Review of Precision Agriculture Technologies and Its Scope of Adoption in India. J. Exp. Agric. Int. 2011, 49–68.

34. Sujatha, R.; Chatterjee, J.M.; Jhanjhi, N.Z.; Brohi, S.N. Performance of Deep Learning vs Machine Learning in Plant Leaf Disease Detection. Microprocess. Microsyst. 2021, 80, 103615.

35. 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.

36. Cavalcante, I.M.; Frazzon, E.M.; Forcellini, F.A.; Ivanov, D. A Supervised Machine Learning Approach to Data-Driven Simulation of Resilient Supplier Selection in Digital Manufacturing. Int. J. Inf. Manag. 2019, 49, 86–97.

37. Arsenovic, M.; Karanovic, M.; Sladojevic, S.; Anderla, A.; Stefanovic, D. Solving Current Limitations of Deep Learning Based Approaches for Plant Disease Detection. Symmetry 2019, 11, 939.

38. Fuentes, A.; Yoon, S.; Kim, S.C.; Park, D.S. A Robust Deep-Learning-Based Detector for Real-Time Tomato Plant Diseases and Pests Recognition. Sensors 2017, 17, 2022.

39. Sperschneider, J. Machine Learning in Plant–Pathogen Interactions: Empowering Biological Predictions from Field Scale to Genome Scale. New Phytol. 2020, 228, 35–41.

40. Adhikari, P.; Araya, H.; Aruna, G.; Balamatti, A.; Banerjee, S.; Baskaran, P.; Barah, B.C.; Behera, D.; Berhe, T.; Boruah, P. System of Crop Intensification for More Productive, Resource-Conserving, Climate-Resilient, and Sustainable Agriculture: Experience with Diverse Crops in Varying Agroecologies. Int. J. Agric. Sustain. 2018, 16, 1–28.

41. De Pascale, S.; Dalla Costa, L.; Vallone, S.; Barbieri, G.; Maggio, A. Increasing Water Use Efficiency in Vegetable Crop Production: From Plant to Irrigation Systems Efficiency. HortTechnology 2011, 21, 301–308.

42. Passioura, J.B.; Angus, J.F. Improving Productivity of Crops in Water-Limited Environments. Adv. Agron. 2010, 106, 37–75.

43. Yang, R.-C.; Crossa, J.; Cornelius, P.L.; Burgueño, J. Biplot Analysis of Genotype× Environment Interaction: Proceed with Caution. Crop Sci. 2009, 49, 1564–1576.

44. Ewert, F.; Rötter, R.P.; Bindi, M.; Webber, H.; Trnka, M.; Kersebaum, K.C.; Olesen, J.E.; van Ittersum, M.K.; Janssen, S.; Rivington, M. Crop Modelling for Integrated Assessment of Risk to Food Production from Climate Change. Environ. Model. Softw. 2015, 72, 287–303.

45. Finger, R.; Schmid, S. Modeling Agricultural Production Risk and the Adaptation to Climate Change. Agric. Finance. 2008, 68, 25–41.

46. Antonopoulos, V.Z.; Antonopoulos, A.V. Daily Reference Evapotranspiration Estimates by Artificial Neural Networks Technique and Empirical Equations Using Limited Input Climate Variables. Comput. Electron. Agric. 2017, 132, 86–96.

47. Kocev, D.; Džeroski, S.; White, M.D.; Newell, G.R.; Griffioen, P. Using Single-and Multi-Target Regression Trees and Ensembles to Model a Compound Index of Vegetation Condition. Ecol. Model. 2009, 220, 1159–1168.

48. Jung, J.; Maeda, M.; Chang, A.; Bhandari, M.; Ashapure, A.; Landivar-Bowles, J. The Potential of Remote Sensing and Artificial Intelligence as Tools to Improve the Resilience of Agriculture Production Systems. Curr. Opin. Biotechnol. 2021, 70, 15–22.

49. Rahaman, M.; Chen, D.; Gillani, Z.; Klukas, C.; Chen, M. Advanced Phenotyping and Phenotype Data Analysis for the Study of Plant Growth and Development. Front. Plant Sci. 2015, 6, 619.

50. Wang, H.; Cimen, E.; Singh, N.; Buckler, E. Deep Learning for Plant Genomics and Crop Improvement. Curr. Opin. Plant Biol. 2020, 54, 34–41.

51. Crane-Droesch, A. Machine Learning Methods for Crop Yield Prediction and Climate Change Impact Assessment in Agriculture. Environ. Res. Lett. 2018, 13, 114003.

52. Basso, B.; Liu, L. Seasonal Crop Yield Forecast: Methods, Applications, and Accuracies. Adv. Agron. 2019, 154, 201–255.

53. Tidake, A.H. Design and Implement a Novel Algorithm to Maximize the Yield of Farming Using Prescriptive Analysis. Available online: http://arxiv.org/abs/2003.00676 (accessed on 30 July 2021).

54. Du, K.-L.; Swamy, M.N. Neural Networks and Statistical Learning; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013.

55. Huang, X.; Jensen, J.R. A Machine-Learning Approach to Automated Knowledge-Base Building for Remote Sensing Image Analysis with GIS Data. Photogramm. Eng. Remote Sens. 1997, 63, 1185–1193.

56. Shakoor, N.; Lee, S.; Mockler, T.C. High Throughput Phenotyping to Accelerate Crop Breeding and Monitoring of Diseases in the Field. Curr. Opin. Plant Biol. 2017, 38, 184–192.

57. Jha, K.; Doshi, A.; Patel, P.; Shah, M. A Comprehensive Review on Automation in Agriculture Using Artificial Intelligence. Artif. Intell. Agric. 2019, 2, 1–12.

58. Padarian, J.; Minasny, B.; McBratney, A.B. Machine Learning and Soil Sciences: A Review Aided by Machine Learning Tools. Soil 2020, 6, 35–52.

59. Kamilaris, A.; Prenafeta-Boldú, F.X. Deep Learning in Agriculture: A Survey. Comput. Electron. Agric. 2018, 147, 70–90.

60. Rehman, T.U.; Mahmud, M.S.; Chang, Y.K.; Jin, J.; Shin, J. Current and Future Applications of Statistical Machine Learning Algorithms for Agricultural Machine Vision Systems. Comput. Electron. Agric. 2019, 156, 585–605.

61. Selvaraj, M.G.; Valderrama, M.; Guzman, D.; Valencia, M.; Ruiz, H.; Acharjee, A. Machine Learning for High-Throughput Field Phenotyping and Image Processing Provides Insight into the Association of above and below-Ground Traits in Cassava (Manihot Esculenta Crantz). Plant Methods 2020, 16, 1–19.

62. Sun, A.Y.; Scanlon, B.R. How Can Big Data and Machine Learning Benefit Environment and Water Management: A Survey of Methods, Applications, and Future Directions. Environ. Res. Lett. 2019, 14, 073001.

63. Virnodkar, S.S.; Pachghare, V.K.; Patil, V.C.; Jha, S.K. Remote Sensing and Machine Learning for Crop Water Stress Determination in Various Crops: A Critical Review. Precis. Agric. 2020, 21, 1121–1155.

64. Maimaitijiang, M.; Sagan, V.; Sidike, P.; Daloye, A.M.; Erkbol, H.; Fritschi, F.B. Crop Monitoring Using Satellite/UAV Data Fusion and Machine Learning. Remote Sens. 2020, 12, 1357.

65. Mir, R.R.; Reynolds, M.; Pinto, F.; Khan, M.A.; Bhat, M.A. High-Throughput Phenotyping for Crop Improvement in the Genomics Era. Plant Sci. 2019, 282, 60–72.

66. Yang, W.; Feng, H.; Zhang, X.; Zhang, J.; Doonan, J.H.; Batchelor, W.D.; Xiong, L.; Yan, J. Crop Phenomics and High-Throughput Phenotyping: Past Decades, Current Challenges, and Future Perspectives. Mol. Plant 2020, 13, 187–214.

67. Adam, E.; Mutanga, O.; Rugege, D. Multispectral and Hyperspectral Remote Sensing for Identification and Mapping of Wetland Vegetation: A Review. Wetl. Ecol. Manag. 2010, 18, 281–296.

68. Govender, M.; Chetty, K.; Bulcock, H. A Review of Hyperspectral Remote Sensing and Its Application in Vegetation and Water Resource Studies. Water Sa 2007, 33, 145–151.

69. Wang, L.; Qu, J.J. Satellite Remote Sensing Applications for Surface Soil Moisture Monitoring: A Review. Front. Earth Sci. China 2009, 3, 237–247.

70. Gonsamo Gosa, A. Remote Sensing of Leaf Area Index: Enhanced Retrieval from Close-Range and Remotely Sensed Optical Observations; Helsingin yliopisto: Helsinki, Finland, 2009.

71. Guyon, D.; Bréda, N. Applications of Multispectral Optical Satellite Imaging in Forestry. In Land Surface Remote Sensing in Agriculture and Forest; Elsevier: Amsterdam, The Netherlands, 2016; pp. 249–329.

72. Alhnaity, B.; Pearson, S.; Leontidis, G.; Kollias, S. Using Deep Learning to Predict Plant Growth and Yield in Greenhouse Environments. In Proceedings of the International Symposium on Advanced Technologies and Management for Innovative Greenhouses: GreenSys2019 1296, Angers, France, 16–20 June 2019; pp. 425–432.

73. Collard, B.C.; Jahufer, M.Z.Z.; Brouwer, J.B.; Pang, E.C.K. An Introduction to Markers, Quantitative Trait Loci (QTL) Mapping and Marker-Assisted Selection for Crop Improvement: The Basic Concepts. Euphytica 2005, 142, 169–196.

74. Bharadwaj, D.N. Advanced Molecular Plant Breeding: Meeting the Challenge of Food Security; CRC Press: Boca Raton, FL, USA, 2018.

75. Liu, W.; Wang, Z.; Liu, X.; Zeng, N.; Liu, Y.; Alsaadi, F.E. A Survey of Deep Neural Network Architectures and Their Applications. Neurocomputing 2017, 234, 11–26.

76. Garrick, D.J.; Fernando, R.L. Implementing a QTL detection study (GWAS) using genomic prediction methodology. In Genome-wide Association Studies and Genomic Prediction; Springer: Berlin/Heidelberg, Germany, 2013; pp. 275–298.

77. Wang, X.; Xu, Y.; Hu, Z.; Xu, C. Genomic Selection Methods for Crop Improvement: Current Status and Prospects. Crop J. 2018, 6, 330–340.

78. Khaki, S.; Wang, L. Crop Yield Prediction Using Deep Neural Networks. Front. Plant Sci. 2019, 10, 621.

79. Acharjee, A.; Kloosterman, B.; Visser, R.G.; Maliepaard, C. Integration of Multi-Omics Data for Prediction of Phenotypic Traits Using Random Forest. BMC Bioinform. 2016, 17, 363–373.

80. Karami, O. Factors Affecting Agrobacterium-Mediated Transformation of Plants. Transgenic Plant J. 2008, 2, 127–137.

81. Sjahril, R.; Mii, M. High-Efficiency Agrobacterium-Mediated Transformation of Phalaenopsis Using Meropenem, a Novel Antibiotic to Eliminate Agrobacterium. J. Hortic. Sci. Biotechnol. 2006, 81, 458–464.

82. Kemppainen, M.J.; Crespo, M.C.A.; Pardo, A.G. Agrobacterium tumefaciens-mediated transformation of ectomycorrhizal fungi. In Diversity and Biotechnology of Ectomycorrhizae; Springer: Berlin/Heidelberg, Germany, 2011; pp. 123–141.

83. Niazian, M.; Niedba\la, G. Machine Learning for Plant Breeding and Biotechnology. Agriculture 2020, 10, 436.

84. Talebi, S.F.; Saharkhiz, M.J.; Kermani, M.J.; Sharafi, Y.; Raouf Fard, F. Effect of Different Antimitotic Agents on Polyploid Induction of Anise Hyssop (Agastache Foeniculum L.). Caryologia 2017, 70, 184–193.

85. Lucchesini, M.; Mensuali-Sodi, A. Plant tissue culture—An opportunity for the production of nutraceuticals. In Bio-Farms for Nutraceuticals; Springer: Berlin/Heidelberg, Germany, 2010; pp. 185–202.

86. Ikeuchi, M.; Ogawa, Y.; Iwase, A.; Sugimoto, K. Plant Regeneration: Cellular Origins and Molecular Mechanisms. Development 2016, 143, 1442–1451.

87. Us-Camas, R.; Rivera-Solís, G.; Duarte-Aké, F.; De-la-Pena, C. In Vitro Culture: An Epigenetic Challenge for Plants. Plant Cell Tissue Organ Cult. 2014, 118, 187–201.

88. Hesami, M.; Jones, A.M.P. Application of Artificial Intelligence Models and Optimization Algorithms in Plant Cell and Tissue Culture. Appl. Microbiol. Biotechnol. 2020, 104, 1–37.

89. Grativol, C.; Hemerly, A.S.; Ferreira, P.C.G. Genetic and Epigenetic Regulation of Stress Responses in Natural Plant Populations. Biochim. Biophys. Acta BBA Gene Regul. Mech. 2012, 1819, 176–185.

90. Mochida, K.; Shinozaki, K. Advances in Omics and Bioinformatics Tools for Systems Analyses of Plant Functions. Plant Cell Physiol. 2011, 52, 2017–2038.

91. Pan, X.; Shen, H.-B. Predicting RNA–Protein Binding Sites and Motifs through Combining Local and Global Deep Convolutional Neural Networks. Bioinformatics 2018, 34, 3427–3436.

92. Libbrecht, M.W.; Noble, W.S. Machine Learning Applications in Genetics and Genomics. Nat. Rev. Genet. 2015, 16, 321–332.

93. Cobb, J.N.; DeClerck, G.; Greenberg, A.; Clark, R.; McCouch, S. Next-Generation Phenotyping: Requirements and Strategies for Enhancing Our Understanding of Genotype–Phenotype Relationships and Its Relevance to Crop Improvement. Theor. Appl. Genet. 2013, 126, 867–887.

94. Civelek, M.; Lusis, A.J. Systems Genetics Approaches to Understand Complex Traits. Nat. Rev. Genet. 2014, 15, 34–48.

95. Wei, G.-W. Protein Structure Prediction beyond AlphaFold. Nat. Mach. Intell. 2019, 1, 336–337.