

<https://doi.org/10.33472/AFJBS.6.9.2024.3275-3286>



African Journal of Biological Sciences

Journal homepage: <http://www.afjbs.com>



Research Paper

Open Access

## Utilizing Plants to Collaborate with Machinery Opportunities in Plant Biology using Machine Learning

Abu Sarwar Zamani<sup>1\*</sup>, Aisha Hassan Abdalla Hashim<sup>2</sup>, Ahmad Talha Siddiqui<sup>3</sup>, Rashid Ayub<sup>4</sup>, Sumbul Alam<sup>5</sup>

<sup>1,2</sup>Department of Electrical and Computer Engineering, International Islamic University Malaysia, Kuala Lumpur 53100, Malaysia.

<sup>1</sup>INTI International University, Persiaran Perdana BBN Putra Nilai, 71800 Nilai, Negeri Sembilan, Malaysia.

<sup>3</sup>Department of CS & IT, Maulana Azad National Urdu University, India.

<sup>4</sup>Department of Science Technology & Innovation, King Saud University, P. O. Box-2454, Riyadh-11451, Saudi Arabia.

<sup>5</sup>Department of Computer Science Engineering, Presidency University, Bengaluru-560064, Karnataka, India.

\*Corresponding: Abu Sarwar Zamani: [sarwar\\_zamani@yahoo.com](mailto:sarwar_zamani@yahoo.com)

Article History

Volume 6, Issue 9, 2024

Received: 11 Apr 2024

Accepted: 04 May 2024

doi: 10.33472/AFJBS.6.9.2024.3275-3286

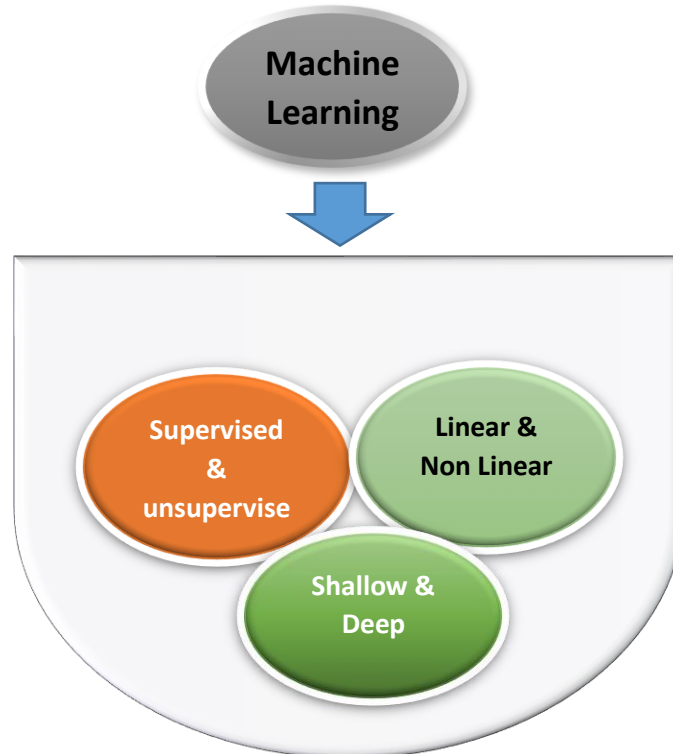
**Abstract:** Machine learning and artificial intelligence are widely used in many scientific and technological domains, and they have lately been used to enhance several plant tissue culture stages. Technological advancements have transformed the way plant genotypes and phenotypes are measured, resulting in the everyday generation of massive, complicated data sets. Specifically, pertinent to the advancement of innovative food and farming technologies is the creation of verified, significant, and practical methods for merging, contrasting, and displaying extensive, multifaceted data sets from various sources and scientific methodologies. Simultaneously, machine learning has advanced quickly, becoming extensively used in science generally and in plant genotyping and phenotyping specifically. In this article, we examine machine learning's use in the field of plant science. We demonstrate how a variety of techniques provided by machine learning allow scientists to identify significant patterns in pertinent plant data.

**Keywords:** Machine Learning, Plant Science, DNN, SNN

## 1. Introduction

Research in plants and agriculture and its application to global social and environmental problems such as food security, biodiversity, and climate change are heavily dependent on data science. The field of artificial intelligence (AI) presents significant promise for controlling and clarifying the complexity of biological systems, organisms, and data. This approach holds great promise for the field of plant sciences, as it addresses the unique challenge of comprehending intricate gene-environment interactions spanning various scales, from the cellular level to the microbiome and climate systems. Additionally, it addresses how these interactions interact with rapidly evolving human management practices in agricultural and other settings, where the reliance on digital innovations is increasing at a rapid pace [1,2]. Consequently, there are more and more examples of helpful AI applications, especially Machine Learning (ML) in plant science settings, and this trend is being further sped up by the Covid-19 epidemic [3,35].

Plant breeding studies, both traditional and in vitro, have made extensive use of machine learning, the discipline of teaching computers to learn from data, to interpret the flow of information about plants from the DNA sequence to the observable phenotypes. Machine learning techniques fall into three categories: shallow, deep, linear, and nonlinear algorithms; supervised, and unsupervised (Figure 1). Artificial neural networks (ANNs), deep neural networks (DNNs), convolutional neural networks (CNNs), random forests (RFs), and support vector machines (SVMs) are nonlinear nonparametric machine learning algorithms used to process nonlinear data in plant research [4]. By utilizing all available spectral data and avoiding multicollinearity and unnecessary spectral bands, these data-driven models may effectively analyze and comprehend unpredictable, non-normal, nonlinear, and nondeterministic data sets [5, 6]. For life science challenges, supervised learning is more effective and appropriate among various learning algorithms, such as reinforcement, unsupervised learning, sparse dictionary, and rule-based learning [7]. Regression (predicting numerical responses) and classification (predicting non-numeric answers) can both be accomplished with supervised learning [8]. Machine learning techniques can be used to evaluate datasets that lack structure, such as those gathered through sequencing or picture imaging [9]. Transposable elements can be identified and categorized using machine learning models that utilize data from genome sequencing [10]. A single model can be created using many inputs mixed together with various combinations of outputs (multiple dependent variables), thus reducing the number of analyses required by breeders.



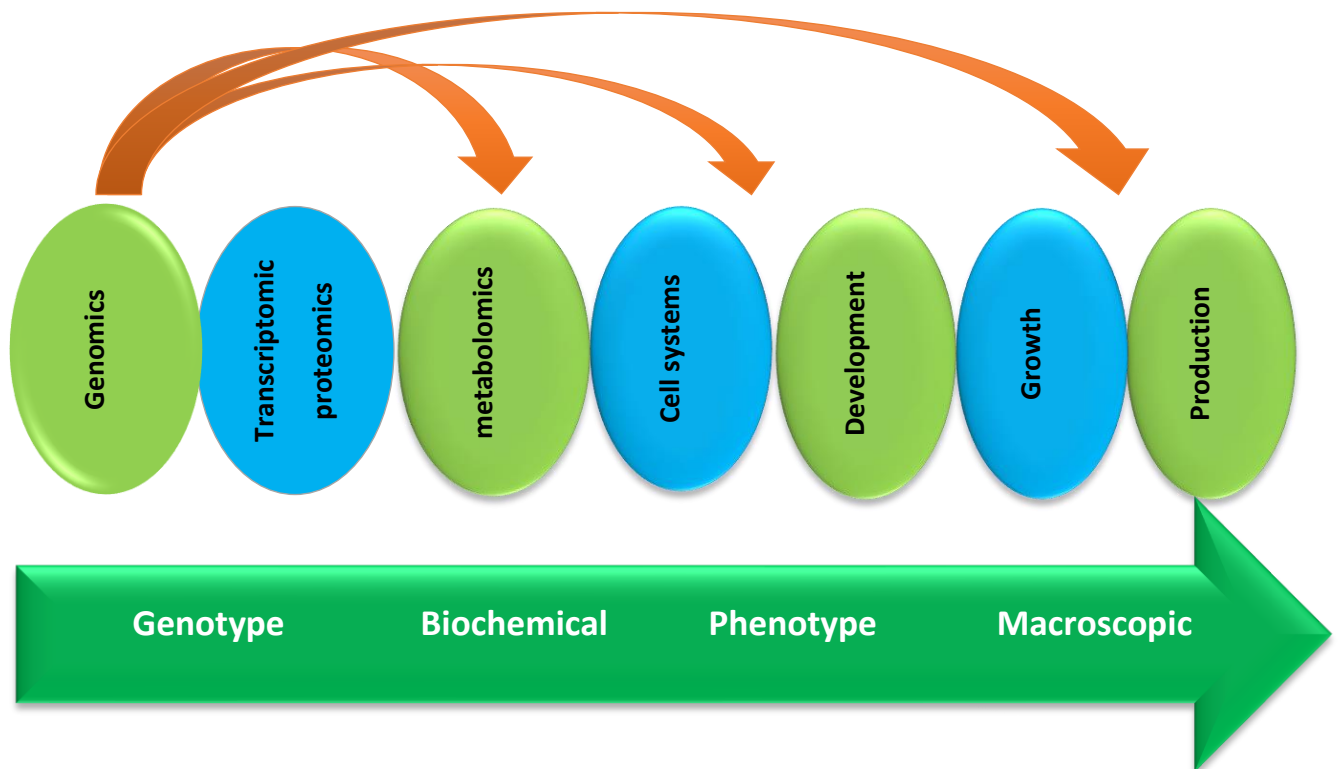
**Figure 1.** Different categories of Machine Learning Algorithm

Genome and genotype analysis is a key component of many modern plant research fields. The full genomes of many plants, including model organisms like *Arabidopsis thaliana*, flowering plant species of scientific interest, trees, algae, and mosses, as well as commercially significant crops like rice, maize, cotton, and wheat, have been determined thanks to the revolution in DNA sequencing. Since genetic variation can be measured with ease in a variety of ways, including single-nucleotide polymorphisms (SNPs), minor insertions and deletions, gene copy number variation, and genome structural variation, population-wide genotyping collections for numerous species are now available [11]. Similar advancements enable the systematic measurement of genome-wide variations at the biochemical level, or in the interactions and concentrations of components in the cell such RNA, proteins, and metabolites, resulting in data sets referred to as "-omics." High-throughput, automated measurements of macroscopic traits can now be used to quantify plant development, morphology, growth, and yield in tissues, organs, or entire plants [12].

Explaining or predicting phenotypes from the underlying genotypes under various environmental situations is a challenge in both basic and practical research (e.g., breeding applications) (Figure 2). Genetic variation results in variations in the biochemical composition of cells, which in turn affects organ formation, plant growth, and ultimately characteristics important to agriculture, such yield and pest and stress tolerance. Modern molecular plant breeding relies on genotype predictive capability to predict yield and quality traits. A fundamental understanding of the regulation of plant development and physiology can be gained by analyzing the effects of genotypic variation and environment on phenotypes. Processing and integrating huge, noisy, and heterogeneous data sets is becoming more and more necessary for analyzing phenotypes obtained

at these many levels or connecting these phenotypes to genotypes. A group of computational techniques called machine learning (ML) are used to identify predictive patterns.

This article discusses the ways in which machine learning (ML) advances the fields of plant science and plant breeding. It focuses on applications in biochemical and macroscopic studies, as well as the relationship between genotypes and these phenotypic levels. Our goal is to demonstrate to non-experts how machine learning (ML) provides a range of techniques that allow us to identify significant patterns in pertinent plant data. Additionally, we offer a critical analysis of ML applications and suggest future areas for research.



**Figure 2.** At different biochemical levels of organization both at the cellular and macroscopic levels, genotype variation is related to phenotypic variation in plant sciences and breeding.

## 2. Literature Review

Machine learning is making significant strides in the field of plant sciences, particularly in the area of automatic species identification from digital herbarium specimen photos [13,14]. This is especially true for genera when there is little to no morphological difference between species, especially when intraspecific taxa are present and hybridization has occurred. In plant science, supervised machine learning techniques are the most often employed techniques. Machine learning techniques are frequently utilized to automate or shorten the time and effort required for researchers to do tasks that were previously accomplished manually. Supervised techniques work

effectively for these kinds of assignments. However, machine learning techniques also offer tools for unsupervised data exploration and data mining [15,36].

To identify plant leaf diseases, a nine-layer deep CNN model was trained using 39 different classes of plant leaves and background images. The resulting 96.46% classification accuracy was found to be higher than that of the more conventional machine learning techniques of SVM, decision trees, logistic regression, and K-NN [16]. CNNs may be used for pattern recognition and object detection in remote sensing applications. An algorithm based on deep CNN-based segmentation that uses visual interpretation of high-resolution Red-Green-Blue (RGB) images based on unmanned aerial vehicles (UAVs) may be used to fine-grained map plant species and communities that may achieve high accuracy (84%) [17].

Because of the enormous data sets, lack of mechanistic knowledge, and requirement for data integration to interpret measurements resulting from data complexity, machine learning (ML) is frequently used to assess findings. Here, we concentrate on recent advancements and, specifically, on situations where the use of ML enables us to look at the part that molecular components play in determining plant phenotypes. We do not discuss machine learning (ML) applications that handle raw measurement data (e.g., in the study of long-read sequencing data [18,37] or well-established methods for analyzing data related to omics [19], such as protein localization or gene models prediction.

One area where machine learning has demonstrated its value is in the differentiation of various genetic regions. For instance, using characteristics like DNA methylation, DNA sequence areas in maize were divided into active genes and (inactive) pseudogenes [20]. Additionally, ML was utilized to forecast the locations of genome crossovers—that is, the areas where genetic material from the mother's and father's genomes is likely to be exchanged—in the genome [21]. As a third illustration, machine learning is beginning to be used in population genetics, but primarily on humans thus far [22]. The identification of genomic areas in plants where natural selection resulted in the nearly total fixing of a particular collection of mutations is one such example [23]. These examples show how machine learning (ML) is currently being used for further research into genome function, in addition to its conventional usage in annotating the structure and function of genes in freshly sequenced genomes [24]. Here, machine learning (ML) is complimentary to more conventional comparative genomics methods with its emphasis on identifying predicted patterns.

Currently, the majority of phenotypic data collection on a macroscopic scale is done manually by human specialists who assess various phenotypes. This drastically limits both the amount and quality of data that is accessible. The comprehension of the relationships between genotype and phenotype is slowed considerably by this phenotyping bottleneck. Digital plant phenotyping, which aims to automatically infer phenotypes from sensor data, has become an active research subject with the goal of accelerating plant science [25].

### 3. Methodology

**3.1 Diversity of data and persistent barriers to data sharing:** When compared to other sciences, biological research is often highly fragmented. Biological attention is paid to the unique features of the target systems they study in order to achieve this. The interaction and behavior of bacteria, fungi, trees, ferns, and mammals with their environment can have a fundamentally different impact on their structures, functions, and reproductive processes due to their fundamental differences in their environment. Numerous additional forms of data lack standardization, and the variety of data formats and methodologies used in various life sciences fields is likely to have an impact on how FAIR principles are used. The ability for machine learning to integrate data across numerous domains may be limited once more by this uneven adoption of FAIR concepts and resources [26].

**3.2 Data selection and digitization:** The plant sciences contain a vast array of data kinds, formats, and sources, making it difficult to choose which data sources to use for AI-informed research. Exist any data sets that, if properly curated, have instant potential? What information is required to characterize data sets so that it is possible to determine whether or not they are appropriate for inclusion in a certain analysis? Achieving well-defined criteria and objectives for data selection is crucial, considering the substantial workload associated in digitizing, curating, and processing datasets and their associated metadata. The criteria must to take into account the machine learning job at hand, the scientific objectives, and the interests of the persons and groups who own the data.

**3.3 Connecting data to materials Samples:** For the results to be interpreted, reused, and reproducible, it is essential to provide clear information on the relationship between digital data and material samples, such as seeds, germplasm, and other biological sources linked to the data [27]. The use of best practices in standards to provide effect through AI/ML is becoming increasingly important, thanks in large part to these efforts and a broader movement in research culture. Companies that create scientific equipment and research software are essential in this regard. This phenomenon is especially noticeable when it comes to data produced by remote sensing technologies, since the predominant standards pertain to the technical aspects of imaging and data processing, rather than data curation.

**3.4 Reliability in data access management:** Datasets that are relevant to biological and environmental studies are required for the use of AI tools in environmental research. However, this is dependent on a number of factors that go well beyond the needs of science, such as intellectual property laws, institutional data governance, and the rights and risks of sharing data for those who generate the materials from which data are extracted and/or may face social and economic repercussions from particular applications of data analysis [28].

### 4. AI-Based Macroscopic Phenotype Detection

Currently, the majority of phenotypic data collection on a macroscopic scale is done manually by human specialists who assess various phenotypes. This drastically limits both the amount and

quality of data that is accessible. The comprehension of the relationships between genotype and phenotype is slowed considerably by this phenotyping bottleneck. Digital plant phenotyping, which aims to automatically infer phenotypes from sensor data, has become an active research subject with the goal of accelerating plant science [29]. In addition to phenotypes of plant organs and reproductive components (development), we are also interested in phenotypes of the whole plant (growth), as well as field (production) phenotypes. Plant characteristics such as leaf area, internode length, root volume, fruit size, chlorophyll content, photosynthetic activity, plant height, biomass, stress on plants, and yield predictions are examples of these characteristics. Certain plant characteristics are typically related to environmental factors such as temperature, light intensity, humidity, soil composition, as well as CO<sub>2</sub> and oxygen concentrations. Traits of plants may be monitored throughout time to investigate phenological features such as growth. Furthermore, by comparing the characteristics of nearby plants, impacts at the field level may be investigated.

This is particularly true when researching intricate plants in noisy, uncontrolled, and chaotic settings. The application of ML in plant phenotyping was promoted [30] as a way to address these complications, especially with regard to picture data. Traditional computer vision systems involve two stages: feature extraction through labor-intensive image processing and machine learning techniques for decision-making. Supervised machine learning techniques are generally employed to extract knowledge from examples, which can reveal patterns in the frequently high-dimensional feature space that are invisible to humans. The drawback is that the handcrafted features' quality determines how well they operate. Recently, end-to-end learning has been introduced in DL approaches that address this restriction. Additionally, DL has been effectively used in plant phenotyping. For example, in order to examine plant growth, [31] proposed the use of a convolutional neural network (CNN) to identify and categorize spikes and spikelets in photographs of wheat.

### **5. Genomic Prediction Using Machine Learning**

Using the available genetic, phenotypic, and environmental data to explain a complex attribute like yield is a primary goal in both plant research and breeding. In order to do this, ML and other methods focus on these aspects. Genomic areas linked to a certain characteristic. This is called genome-wide association mapping (or QTL mapping for experimental populations). Calculate the contributions of each locus individually and the percentage of trait variance accounted for by the collective loci. In connection with this, genetic correlations between characteristics are interesting since they measure how much genetic signals overlap. Forecast the anticipated trait values for novel genotypes (genomic prediction [GP]), for which only marker data are provided. The use of genomic selection [32], which selects material based on GPs rather than phenotypic characteristics, and marker-assisted selection, which requires QTL mapping in order to get desirable alleles at certain loci, is becoming more and more common among plant breeders. Genetic architecture and the genes that underlie QTLs are of primary interest to biologists.

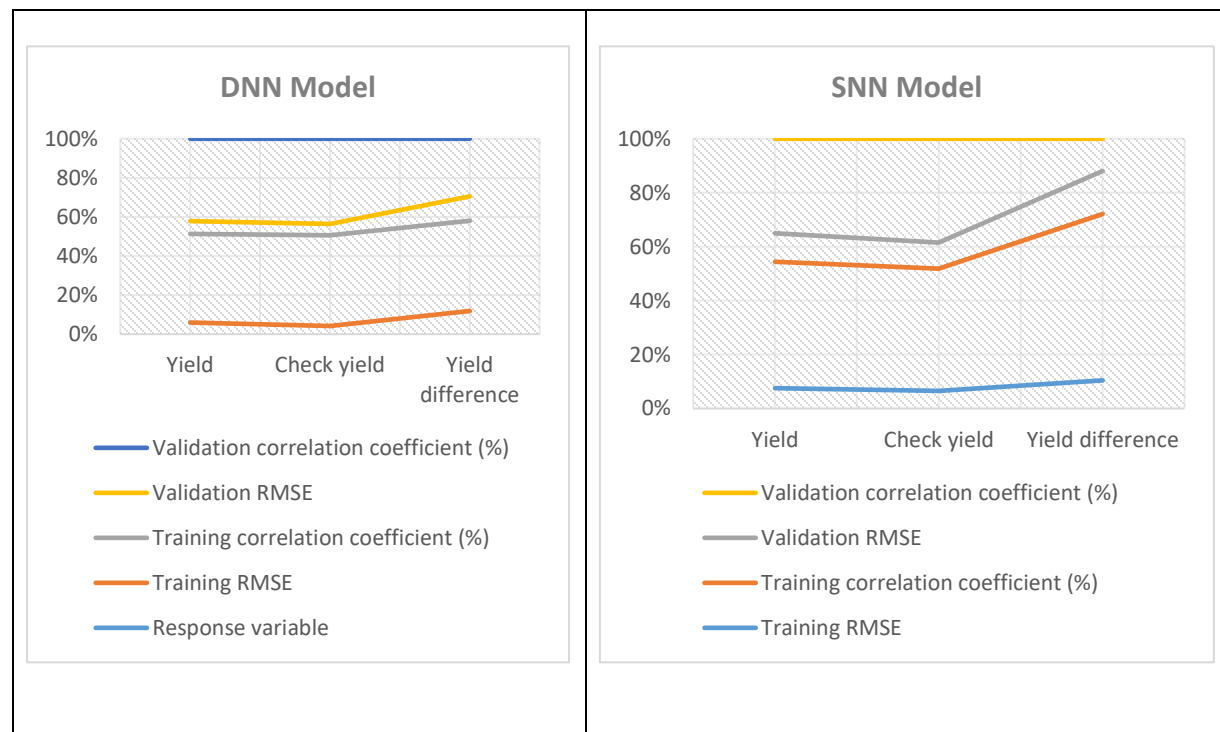
At least in comparison to animal or human genetics, the limited sample size presents a first obstacle in the use of machine learning here. With a few notable exceptions, genotypes usually

number in the hundreds, despite the fact that phenotypes themselves can be quite high dimensional. However, a successful implementation of machine learning will frequently require bigger populations due to the vast number of species, genotypes, and habitats researched in the field of plant sciences. This in turn will necessitate various sets of genotypes assessed in separate studies, given the capability of most platforms. Thus, advancements in machine learning in this area will also depend on certain non-ML-related problems, such as enhanced environmental control and measurement precision, higher capacity, and sound experimental design [33].

We used the following models to get the yield prediction data and then compared the relative impact of genotype, soil, and weather components. DNN: This method uses the DNN model to predict the phenotype by analyzing genotype data without considering environment data. The DNN model captures both linear and nonlinear effects of genetic markers. SNN: A SNN model is used to predict the phenotype based on soil data (without having to know genotypes and weather data) for a model that includes both linear and nonlinear impacts of soil conditions.

Model	Response variable	Training RMSE	Training correlation coefficient (%)	Validation RMSE	Validation correlation coefficient (%)
DNN	Yield	10.55	88.3	12.79	81.91
	Check yield	8.21	91.00	11.38	85.46
	Yield difference	11.79	45.87	12.40	29.28
SNN	Yield	12.96	80.21	18.04	60.11
	Check yield	10.24	71.18	15.18	60.48
	Yield difference	9.92	58.74	15.19	11.39

**Table 1.** Prediction accuracy using meteorological ground truth variables [34].



**Figure 2.** The probability density distributions of the yield predicted by the DNN and SNN models and the ground truth yield.



## 6. Conclusions

Our machine learning method for predicting crop (plant science) output was introduced, and it performed exceptionally well when tested on large datasets of corn hybrids during the 2018 Syngenta Crop Challenge. To predict yield (including yield, check yield, and yield difference), deep neural networks (ML) were used based on genetic and environmental data. From historical data, the meticulously crafted deep neural networks were able to learn the intricate and nonlinear relationships between genes, environmental factors, and their interactions. As a result, they were able to reasonably accurately predict the yields of new hybrids planted in areas with known weather patterns. A relatively sensitive relationship between performances of the model and weather forecast accuracy indicates that weather prediction approaches are important. In the future, machine learning (ML) may be used to assist decision-making in a variety of plant research domains, such as anticipating which regions of the genome should be altered to produce a desired phenotype or guaranteeing ideal local growth conditions through the measurement of crop performance *in vivo* in a greenhouse or on the field. Even though these are essentially engineering problems, effective machine learning techniques will provide researchers with strong tools, especially as they improve to the point where their choices can be more easily interpreted. ML can thus assist in addressing the difficulties we encounter in providing food security for expanding people in quickly changing surroundings.

## References:

1. Wang H, Cimen E, Singh N, et al.: Deep Learning for Plant Genomics and Crop Improvements. *Curr Opin Plant Biol.* 2020; 54: 34–41.
2. Harfouche AL, Jacobson DA, Kainer D, et al.: Accelerating climate resilient plant breeding by applying nextgeneration artificial intelligence. *Trends Biotechnol.* 2019; 37(11):1217–1235.
3. King B: Inaugural Address. CGIAR Big Data Convention. 2020; 2020.
4. 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 UsingcMultispectral Images from an Unmanned Aerial Vehicle. *Remote Sens.* 2018, 10, 2026.
5. 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.
6. 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.
7. Asefpour Vakilian, K. Machine learning improves our knowledge about miRNA functions towards plant abiotic stresses. *Sci. Rep.* 2020, 10, 3041.
8. Wang, H.; Cimen, E.; Singh, N.; Buckler, E. Deep learning for plant genomics and crop improvement. *Curr. Opin. Plant Biol.* 2020, 54, 34–41.

9. Hu, H.; Scheben, A.; Edwards, D. Advances in Integrating Genomics and Bioinformatics in the Plant Breeding Pipeline. *Agriculture* 2018, 8, 75.
10. 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.
11. Torkamaneh, D., Boyle, B., and Belzile, F. (2018). “Efficient genome-wide genotyping strategies and data integration in crop plants.” *TAG. Theoretical and applied genetics. Theor. Appl.enet.* 131, 499–511.
12. Zhao, C., Zhang, Y., Du, J., Guo, X., Wen, W., Gu,S., Wang, J., and Fan, J. (2019). Crop phenomics:current status and Perspectives. *Front. Plant Sci.*10, 714.
13. Schuettpelz, E., P. B. Frandsen, R. B. Dikow, A. Brown, S. Orli, M. Peters, A.Metallo, et al. 2017. Applications of deep convolutional neural networks to digitized natural history collections. *Biodiversity Data Journal* 5: e21139.
14. Wäldchen, J., and P. Mäder. 2018. Machine learning for image based species identification.*Methods in Ecology and Evolution* 9: 2216–2225.
15. Soltis, P. S., G. Nelson, A. Zare, and E. K. Meineke. 2020. Plants meet machines: Prospects in machine learning for plant biology. *Applications in Plant Sciences* 8(6): e11371.
16. 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.
17. Kattenborn, T.; Eichel, J.; Wisser, S.; Burrows, L.; Fassnacht, F.E.; Schmidlein, S. Convolutional Neural Networks accurately predict cover fractions of plant species and communities in Unmanned Aerial Vehicle imagery. *Remote Sens. Ecol. Conserv.* 2020.
18. Amarasinghe, S.L., Su, S., Dong, X., Zappia, L., Ritchie, M.E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21, 30.
19. Mahood, E.H., Kruse, L.H., and Moghe, G.D.(2020). Machine learning: a powerful tool for gene function prediction in plants. *Appl. Plant Sci.* 8,e11376.
20. Sartor, R.C., Noshay, J., Springer, N.M., and Briggs, S.P. (2019). Identification of the expressome by machine learning on omics data.*Proc. Natl. Acad. Sci. U S A* 116, 18119–18125, <https://doi.org/10.1073/pnas.1813645116>.
21. Demirci, S., Peters, S.A., de Ridder, D., and van Dijk, A.D.J. (2018). DNA sequence and shape are predictive for meiotic crossovers throughout the plant kingdom. *Plant J.* <https://doi.org/10.1111/tpj.13979>.
22. Schrider, D.R., and Kern, A.D. (2018). Supervised machine learning for population genetics: a new paradigm. *Trends Genet.* 34, 301–312, <https://doi.org/10.1016/j.tig.2017.12.005>.
23. Bourgeois, Y., Stritt, C., Walser, J.-C., Gordon, S.P., Vogel, J.P., and Roulin, A.C. (2018). Genome-wide scans of selection highlight the impact of biotic and abiotic constraints in natural populations of the model grass *Brachypodium distachyon*. *Plant J. Cell Mol. Biol.* 96, 438–451, <https://doi.org/10.1111/tpj.14042>.

24. Yip, K.Y., Cheng, C., and Gerstein, M. (2013). Machine learning and genome annotation: a match meant to be? *Genome Biol.* 14, 1–10. BioMed Central. <https://doi.org/10.1186/gb-2013-14-5-205>.
25. Roitsch, T., Cabrera-Bosquet, L., Fournier, A., Ghamkhar, K., Jimenez-Berni, J., Pinto, F., and Ober, E.S. (2019). Review: new sensors and data driven approaches—a path to next generation phenomics. *Plant Sci.* 282, 2–10, <https://doi.org/10.1016/j.plantsci.2019.01.011>.
26. Pommier C, Michotey C, Cornut G, et al.: Applying FAIR Principles to Plant Phenotypic Data Management in GnpIS. *Plant Phenomics*. 2019;1671403.
27. Strasser B: *Collecting Experiments: Making Big Data Biology*. Chicago:Chicago University Press; 2019.
28. Williamson HF, Leonelli S (eds.): *Towards Responsible Plant Data Linkage: Data Challenges for Agricultural Research and Development*. Springer Open Access; 2022.
29. Roitsch, T., Cabrera-Bosquet, L., Fournier, A., Ghamkhar, K., Jimenez-Berni, J., Pinto, F., and Ober, E.S. (2019). Review: new sensors and data driven approaches—a path to next generation phenomics. *Plant Sci.* 282, 2–10, <https://doi.org/10.1016/j.plantsci.2019.01.011>.
30. Kamilaris, A., and Prenafeta-Boldu´, F.X. (2018). Deep learning in agriculture: a survey. *Comput. Electronics Agric.* 70–90, <https://doi.org/10.1016/j.compag.2018.02.016>.
31. Shi, W., van de Zedde, R., Jiang, H., and Kootstra, G. (2019). Plant-part segmentation using deep learning and multi-view vision. *Biosyst. Eng.* 187, 81–95, <https://doi.org/10.1016/j.biosystemseng.2019.08.014>.
32. Crossa, J., Pe´ rez-Rodr´ ıguez, P., Cuevas, J., Montesinos-Lo´ pez, O., Jarqu´ ın, D., de Los Campos, G., Burguen˜ o, J., Gonz´ alez-Camacho, J.M., Pe´ rez-Elizalde, S., Beyene, Y., Dreisigacker, S., et al. (2017). Genomic selection in plant breeding: methods, models, and Perspectives. *Trends Plant Sci.* 22, 961–975.
33. Selby, P., Abbeloos, R., Backlund, J.E., Basterrechea Salido, M., Bauchet, G., Benites-Alfaro, O.E., Birkett, C., Calaminos, V.C., Carceller, P., Cornut, G., et al. (2019). BrAPI—an application programming interface for plant breeding applications. *Bioinformatics* 35, 4147–4155, <https://doi.org/10.1093/bioinformatics/btz190>.
34. Khaki, Saeed, and Lizhi Wang. "Crop yield prediction using deep neural networks." *Frontiers in plant science* 10 (2019): 452963.
35. Gupta, S., Geetha, A., Sankaran, K. S., Zamani, A. S., Ritonga, M., Raj, R., ... & Mohammed, H. S. (2022). Machine learning-and feature selection-enabled framework for accurate crop yield prediction. *Journal of Food Quality*, 2022, 1-7.
36. Abuzanounh, K. I. M., Al-Wesabi, F. N., Albraikan, A. A., Al Duhayyim, M., Al-Shabi, M., Hilal, A. M., ... & Muthulakshmi, K. (2022). Design of machine learning based smart irrigation System for precision Agriculture. *Computers, Materials & Continua*, 72(1), 109-124.

37. Almasoud, A. S., Abdelmaboud, A., Elfadil Eisa, T. A., Al Duhayyim, M., Hassan Elnour, A. A., Ahmed Hamza, M., ... & Sarwar Zamani, A. (2022). Artificial Intelligence-Based Fusion Model for Paddy Leaf Disease Detection and Classification. *Computers, Materials & Continua*, 72(1).