

# AI and Crop Improvement

true

June 2022

## Abstract

Lorem ipsum dolor sit amet.

## Contents

<b>Crop Improvement</b>	<b>1</b>
Prior Progress . . . . .	1
Modern Crop Improvement . . . . .	2
<b>AI</b>	<b>2</b>
AI in Crop Improvements . . . . .	3
<b>References</b>	<b>3</b>

## Crop Improvement

To achieve a rate of yield increase that keeps pace with the growing world population, new approaches to crop improvement will be required.

### Prior Progress

When comparing crop cultivars of a century ago and those of today, considerable differences are often observed. Improvements to yield, plant architecture, and biotic and abiotic resistances, among other traits, have been targeted by breeding and management strategies since the early part of the 20th century. Through the 1930s to 2000s, both breeding and cultural practices contributed considerably to increases in productivity as measured by average yield increases (Duvick 1999; Duvick 2005).

Conventional breeding took advantage of traditional phenotyping methods to inform crossing and selection.

With the increase in genetic technology availability and decrease in cost during the (1980s-2000s), targeted improvements through approaches such as marker assisted selection became feasible. Genetic improvement was seen as the new way forward in contrast to the plateauing gains made from cultural practices. Genetic data became abundant in the next-generation sequencing era. Earlier QTL studies have identified many loci with large contributions to highly heritable traits. Traits with unresolved genetic architecture, result from the contributions of many, small effect loci. QTL studies in such traits require extensive phenotyping and are often not stable across environment.

The “phenotyping bottleneck” issue has driven increased research on high throughput phenotyping (HTP) platforms and spurred in the generation of “big data” in many facets of crop improvement schemes.

## Modern Crop Improvement

### AI

In recent years, artificial intelligence (AI) has been increasingly explored as a means to analyze big data. Within crop improvement in the 2020s, the processes of genotyping, phenotyping, and envirotyping a single population can produce dramatically more data than was cumulatively obtained in the breeding process in previous decades. While bottlenecks of traditional methods have been overcome with higher levels of automation achieved with post-NGS sequencing technologies, HTP, and remote sensing, fully exploiting these data in an efficient manner is difficult with prior approaches. AI is most commonly applied through machine learning which includes fields like computer vision, neural networks, deep learning, and others. Certain subfields, like those aforementioned, are more commonly referenced than others in conversations revolving around AI and machine learning. However, there is a great deal of overlap between statistical and machine learning Fig. 1. Any approach that makes a prediction via computation can be considered to be machine learning. For example, under this definition, linear models can be considered to be machine learning despite their relative simplicity compared to things like NN and DL.

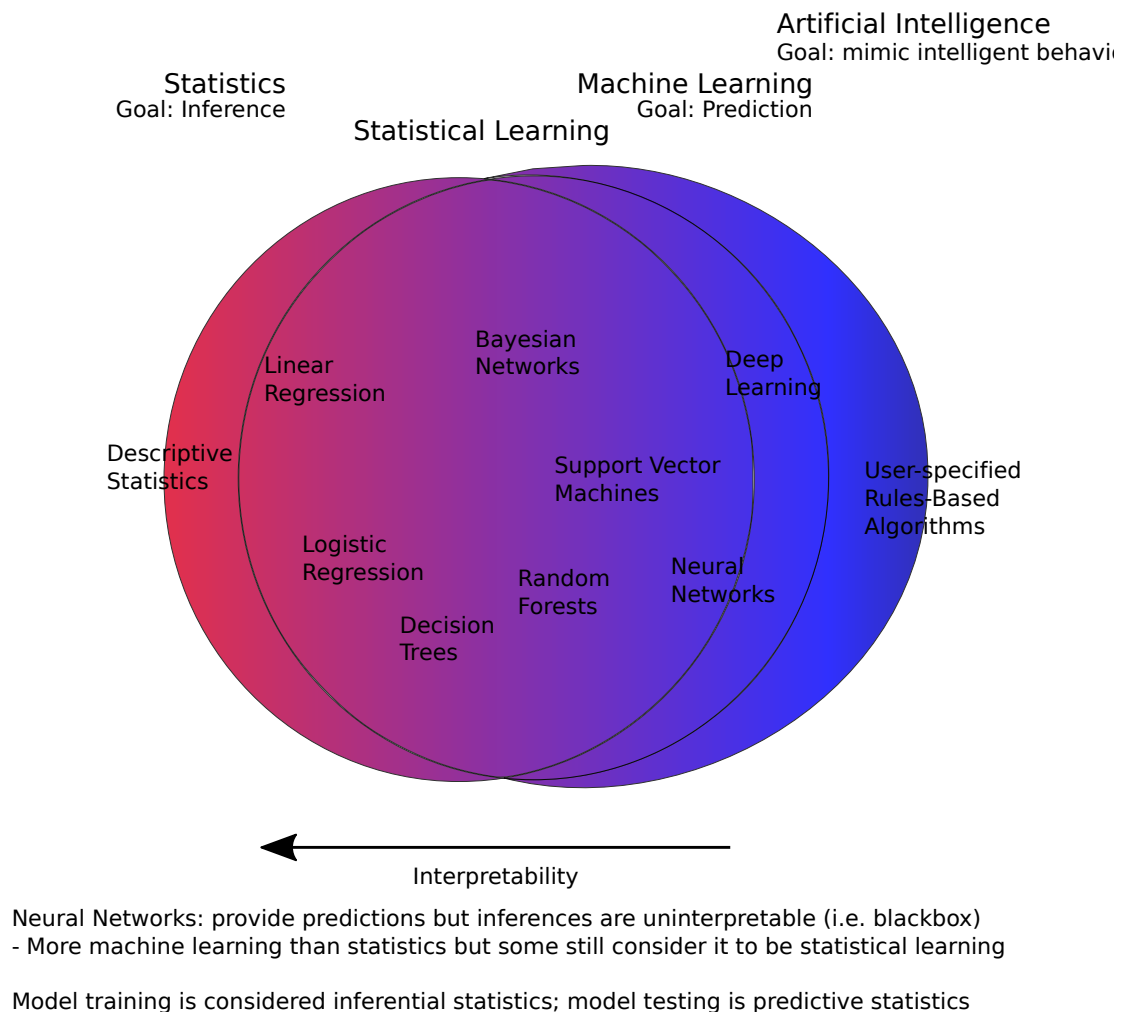


Figure 1: Caption.

Artificial intelligence (AI) is concerned with the process of designing computers that can think and act humanly and rationally (Russell and Norvig 2009). AI algorithms that do not attempt learning are considered

to be rules-based. Rules-based systems are generally implemented by defining a series of situation-action (i.e. if-then) rules. Defining these rules requires previous knowledge of the relationship between input-output pairs (Hayes-Roth 1985). Programming a system that can anticipate all situation-actions which lead from input to output variables is more difficult than training a system that can learn from input (Jordan and Mitchell 2015). At the core of modern AI is machine learning (ML). Machine learning is a subfield within AI which attempts *learning* or improvement through experience (Russell and Norvig 2009; Libbrecht and Noble 2015; Liakos et al. 2018). ML can be characterized by feedback type, data type . . . .

Machine learning can be considered to be supervised or unsupervised given the system’s access to feedback. Supervised learning occurs in two phases, training and testing. During the training phase, predetermined input-response pairs (labeled data) are used as examples and the learning algorithm attempts to formulate function that connect input data to respective labels (Liakos et al. 2018; Montesinos López et al. 2022, p. 29). During the testing phase of supervised learning, the learned pattern (trained model) is used to generate label predictions and the accuracy of the predictions can be evaluated against user-defined labels. In this type of system the feedback is considered to be explicit. In unsupervised learning, the absence of preassigned labels for the input data only allows the model to evaluate patterns and prevents prediction accuracy from being evaluated since no correct input-response pairs have been specified. This type of system has no feedback. Systems that utilize both supervised and unsupervised learning are termed semi-supervised.

A third type of system where feedback is considered to be implicit is termed reinforcement learning. Reinforcement learning

Within unsupervised and supervised ML Machine learning exists at the intersection of computer science and statistics and thus many statistical models are considered ML when applied with a prediction-centric approach. The field of statistics by comparison puts a greater focus on inference over prediction (Bzdok et al. 2018), but methods shared by the two fields can generally accomplish both. For this reason

## AI in Crop Improvements

Prior issues facing crop improvement research revolved around data accumulation. Genotyping methods prior to NGS were laborious and limited quantities of genotypes were achievable. Traditional phenotyping methods are time intensive and sometimes limited by the expertise of evaluators. Envirotyping was difficult to achieve with useful levels of resolution. Considerable progress has been made on these problems by the early 2020s. The post-NGS sequencing technology era has made sequencing whole genomes a feasible genotyping approach. image-based phenotyping and other high throughput techniques generate terabytes of data every growing season. Remote sensing technologies have made observing envirotypes within a single field possible. **transition** However these solutions lend new issues to the field of crop improvement. New issues revolve around the need for approaches which effectively and efficiently analyzing resulting big data; bridging the gap between observed genotypes and observed phenotypes a) yield results that are meaningful biologically (interpretable), b) feasible computationally (efficient), c) have high fidelity to plants represented (accuracy), d) achieves improvement (predictive/useful). **transition**

Achieving improvement while utilizing genomics, phenomics and enviromics requires an integrative approach. A large number of studies have employed genomic selection for this purpose.

## References

- Bzdok D, Altman N, Krzywinski M (2018) Statistics versus machine learning. *Nature Methods* 15:233–234. <https://doi.org/10.1038/nmeth.4642>
- Duvick DN (1999) Heterosis: feeding people and protecting natural resources. *Genetics and exploitation of heterosis in crops* 19–29
- Duvick DN (2005) *The Contribution of Breeding to Yield Advances in maize (Zea mays L.)*. In: *Advances in agronomy*. Academic Press, pp 83–145
- Hayes-Roth F (1985) Rule-based systems. *Communications of the ACM* 28:921–932
- Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. *Science* 349:255–60. <https://doi.org/10.1126/science.aaa8415>

- Liakos K, Busato P, Moshou D, et al (2018) Machine Learning in Agriculture: A Review. Sensors 18:2674. <https://doi.org/10.3390/s18082674>
- Libbrecht MW, Noble WS (2015) Machine learning applications in genetics and genomics. Nature Reviews Genetics 16:321–332. <https://doi.org/10.1038/nrg3920>
- Montesinos López OA, Montesinos López A, Crossa J (2022) **Multivariate Statistical Machine Learning Methods for Genomic Prediction**. Springer International Publishing, Cham
- Russell S, Norvig P (2009) Artificial Intelligence: A Modern Approach, 3rd edn. Prentice Hall Press, USA