Machine Learning Speeding Up the Development of Portfolio of New Crop Varieties to Adapt to and Mitigate Climate Change

Abdallah Bari*
OperAI
Montreal, Canada

Hassan Ouabbou
INRA
Rabat, Morocco

Abderrazek Jilal
INRA
Rabat, Morocco

Hamid Khazaei
World Vegetable Center
Tainan, Taiwan

Frederick L. Stoddard
University of Helsinki
Helsinki, Finland

Mikko J. Sillanpää
University of Oulu
Oulu, Finland

Abstract

Climate change poses a serious challenge to agriculture sector requiring to produce more food to meet the increase in food demand while adapting to and mitigating changing climate conditions. There is an urgent need to speed up the development of new portfolio of crop varieties with traits of adaptation and mitigation to climate change. We have used mathematical approaches, including machine learning approaches (ML), leading to unprecedented results as some of the traits, which have been long sought-for, have been found within a short period of time.

1 Climate change - a dual challenge

Climate change poses a serious challenge to achieving food security. It is a dual challenge that requires keeping up with the world's increasing demand for food, while adapting to and mitigating climate change. The agriculture sector is in the midst of climate change, in a time of a need to keep pace and produce more food to close the gap of 56 percent between the amount of food available today and that required by the year 2050 (WRI 2021).

1.1 Why the need for speed

The average yield of maize crop in Iowa State (USA) during the year 2012, which one of the driest year in a half century, is equivalent to the average yield of the 1980s known for more favorable conditions for plant growth. The genetic gain that was made during the period of 30 years of crop improvement is considerable (Atlin et al. 2017). However, With current frequent dry spells, there is a need to speed up further the rate of the genetic gain. To speed plant improvement Koo and Wright (2000) found that early identification of valuable crop traits is of equal importance to the process of incorporating these traits into an improved genetic background.

To help speed up the development of new crops varieties with traits to adapt to and mitigate changing climate conditions, we have used mathematical approaches, including machine learning techniques. Mathematical approaches have played a major role in producing more food, prior to the emergence of molecular approaches. Fisher (1930) elaborated the mathematical theoretical framework that was considered as the basis of quantitative genetic theory, leading to impressive gains in yields (Scheffé 1959).

*Operational AI to speed up and streamline operational processes at the edge of the Cloud. a.bari@operai.ca

35th Conference on Neural Information Processing Systems (NeurIPS 2021)
.

1.2 Why portfolio of varieties

There is currently an interest in the use of mathematical and machine learning approaches to accelerate and optimize further crop improvement processes to develop new portfolio of diverse crop varieties with enhanced traits (Anderssen and Edwards 2012, Bari et al. 2016, Parmley et al. 2019, Tong and Nikoloski 2021). Diversified systems have been reported to raise productivity and improve livelihoods, performing particularly well under environmental stress and delivering production increases in places where additional food is desperately needed (De Schutter and Frison 2017).

2 Modeling - Evolutionary processes

To mimic evolutionary processes that drive traits natural adaptation in plants we opted for inverse problem modeling involving ML. The inverse problem using actual field observations to estimate values that are not easily directly observed in the field, such as heat and drought traits. We used different machine-learning techniques to accelerate the search for these traits. The techniques span supervised and unsupervised techniques, including Bayes, Neural Network (NN), Random Forest and K-means techniques, to help in the rapid identification of adaptive traits (Cherkassky and Mulier 2007, Khazaei et al. 2013, Bari et al. 2016, Bari 2018). The ML-based search for these adaptive traits is based on exploring and exploiting the dependence between the desired traits (denoted Y) and the environment (denoted X) as an evolutionary co-driver prevailing in the areas where the samples were originally sampled (Henry and Nevo 2014). The performance of ML models was assessed using Kappa and AUC values (Fawcett 2006). The desired traits are considered as representative variables with the additive influence of many genes with small effects (Brown et al. 1996). Crop simulation models combined with high-resolution climate change map scenarios can help to identify key traits that are important under drought and high temperature stresses in crops (Semenov and Halford 2009).

2.1 Traits of adaptation to drought

To identify traits of adaptation to drought, we used the faba bean (*Vicia faba* L.) crop, which is a widely grown food legume crop in the dry areas and is being considered as one of the most likely crop to be impacted by climate change (Duc et al. 2011). Cumulative plant datasets from experiments on faba bean were used to explore the link between the trait expression (Y) and the environment (X). A total of 400 plant samples was originally selected on the basis of extreme "wet" and "dry" environmental profiles using a clustering algorithm from thousands of samples stored in the genebank. From these 400, a first subset was used to detect the presence of patterns, if any, in the data form the training set, and the patterns or dependence detected together with a new X was then used in turn to assign values (*a priori*), as predictive probabilities of having the sought-after drought traits, to another unknown subset (*a posteriori*). The plants of this latter subset were grown for evaluation, beside the first subset, to test the predictions for their accuracy and agreement with the actual field evaluation data. The traits plant data consists of a set of leaf morpho-physiological measurements that capture drought-adaptation related traits used previously (Khazaei et al. 2013). These measurements span plant gas exchange properties, photosynthesis and phenology.

2.2 Traits of adaptation to heat

Here we used unsupervised learning to cluster data using barley (*Hordeum vulgare* L.) crop as the test crop plant species. Clustering was conducted using environmental data (*a priori*) to develop one subset that was likely to contain climate-change–related traits and another subset representative (core) of the different environments where the barley samples were originally sampled across Morocco. The samples allocated to any one cluster shared phenotypic affinity vis a vis either presence or absence of the traits of tolerance to heat. Each subset contained 100 samples, of which 30 samples were selected at random. All the subsets were grown in the same field for comparison based on *a posteriori* evaluation (Jilal et al. 2016).

To compare and validate the sub-setting based on climate data, the samples were compared based on their evaluation attributes. Both subsets were grown in single rows of 5 m length. Observations were taken on several morphological and physiological traits. Further evaluation was carried out using a honeycomb design of hill plots with the aim to also reduce the area needed for evaluation for large number of samples.

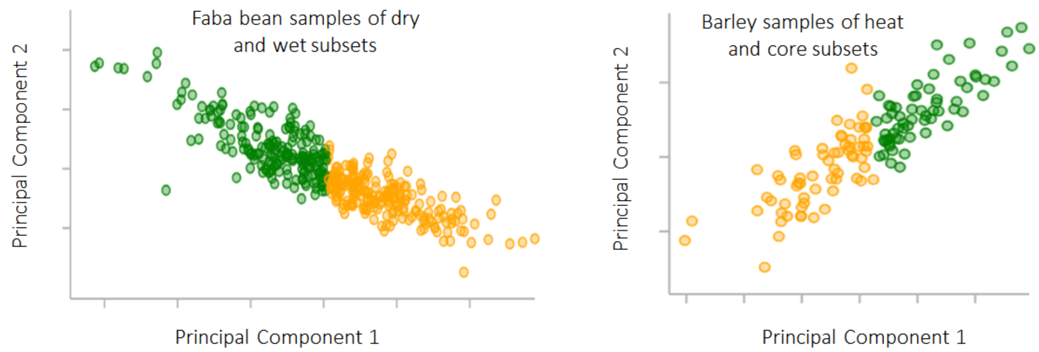

Figure 1: *A posteriori* field evaluation of plants selected by the models (faba bean left, barley right)

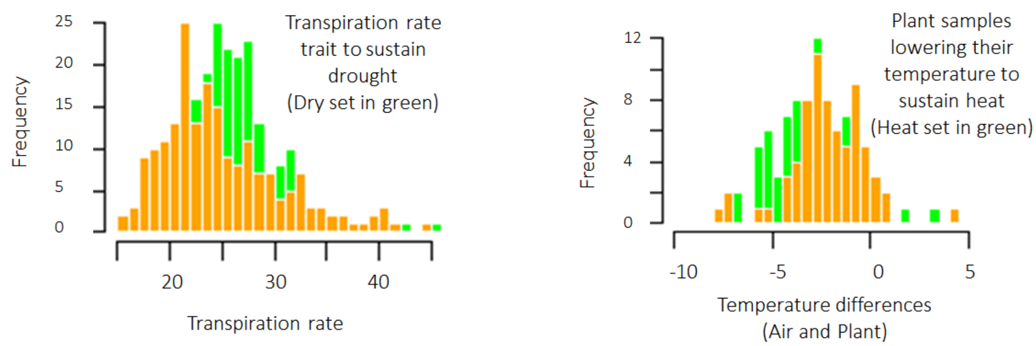

Figure 2: Distrubtions of some of the traits of relevance to drought (left) and heat (right)

3 Results - Traits identified

There was strong agreement between the prediction *a priori* and the actual *a posteriori* field evaluation of the plants in terms of their capability to tolerate drought and heat. Both AUC and Kappa metric values were high, indicating that it is highly likely to identify traits that will provide stress tolerance to crops and can be transferred to cultivars by breeding. The *a posteriori* evaluation showed that the subsets were different (Figure 1). The barley heat subset tended to have the capability to lower its leaf temperature to sustain heat than the core subset while the faba bean dry subset tended to increase it transpiration rate to sustain drough (Figure 2). These results indicate that natural genetic variation contains the much-needed trait variation for adaptation of crops.

The results show that machine learning can also help in reducing field evaluation, as testing can be focused on samples that are highly likely to have the desired traits rather than having to screen the whole samples, which is practically impossible, given their large number. This can help in speeding up the development of portfolio of new crop varieties, while minimizing costs.

4 Conclusion

Mathematical models including machine-learning models can help enormously in identifying the desired traits while shortening the time and reducing significantly the costs to develop new and diverse crop varieties. Machine learning can help to reduce costs that may be incurred to assess and evaluate large number of samples in genebanks. There are more than 1750 genebanks worldwide, holding together more than 7 million of plant samples. ML has the potential to identify rapidly traits, including climate-related traits and to speed up crop development processes to develop portfolios of new crops varieties to adapt to and mitigate climate change.

References

Anderssen, R.S. and Edwards, M.P. (2012) Mathematical modelling in the science and technology of plant breeding. *Int. J. Numer. Anal. Model.* Series B 3:242–258.

Atlin, G.N., Cairns, J.E. and Das, B. (2017) Rapid breeding and varietal replacement are critical to adaptation of cropping systems in the developing world to climate change. *Glob Food Sec.* 12:31-37.

Bari, A. (2018) *Machine Learning at Work: Speeding Data Discovery*. OperAI, Montreal, Canada.

Bari, A., Khazaei, H., Stoddard, F.L., Street, K., Sillanpää, M.J., Chaubey, Y.P. Dayanandan, S. Endresen, D.F., De Pauw, E. and Damania, A.B. (2016) In silico evaluation of plant genetic resources to search for traits for adaptation to climate change. *Climatic Change* 134(4): 667-680.

Bari, A., Damania, A.D., Mackay, M. and Dayanandan, S. (Eds.) (2016) *Applied Mathematics and Omics to Assess Crop Genetic Resources for Climate Change Adaptive Traits* . CRC Press, Taylor and Francis Group, Boca Raton, FL, USA.

Brown, J.H., Stevens, G.C., Kaufman, D.M. (1996) The geographic range: size, shape, boundaries, and internal structure.*Annu. Rev. Ecol. Syst.* 27, 597–623.

Cherkassky, V. and Mulier, F. (2007) *Learning from Data: Concepts, Theory, and Methods*. Wiley-IEEE Press, New York, USA.

De Schutter, O. and Frison, E. (2017) Modern agriculture cultivates climate change – we must nurture biodiversity. *Grain*. Saskatchewan, Canada.

Duc, G., Link, W., Marget, P., Redden, R.J., Stoddard, F.L., Torres, A.M., Cubero, J.I. (2011) Genetic adjustment to changing climates: faba bean. In: Yadav, S.S., Redden, R.J., Hatfield, J.L., Lotze-Campen, H., Hall, A.E. (Eds.). *Crop adaption to climate change* , 1rd ed. John Wiley and Sons, pp. 269–286.

Fawcett, T. (2006) An introduction to ROC analysis. *Pattern Recogn Lett* 27:861–874.

Fisher, R. (1930) Inverse Probability. *Mathematical Proceedings of the Cambridge Philosophical Society* 26(4): 528–535.

Henry, R.J. and Nevo, E. (2014) Exploring natural selection to guide breeding for agriculture. *Plant Biotechnol J* 12: 655–662.

Jilal, A., Ouabbou, H. and Maatougui, M. (2016) Barley Genetic Resources for Climate-Change Adaptation: Searching for Heat-Tolerant Traits through Rapid Evaluation of Subsets. *In: Bari, A., Damania,A.B., Mackay, M. and Dayanandan, S. (Eds.). Applied Mathematics and Omics to Assess Crop Genetic Resources for Climate Change Adaptive Traits.* CRC Press, Taylor and Francis Group, Boca Raton, FL, USA.

Khazaei, H., Street, K., Bari, A., Mackay, M., Stoddard, F.L. (2013) The FIGS (focused identification of germplasm strategy) approach identifies traits related to drought adaptation in *Vicia faba* genetic resources. *PLoS One* 8(5): e63107.

Koo B. and Wright B.D. (2000) The optimal timing of evaluation of genebank accessions and the effects of biotechnology. *Am J Agric Econ* 82:797–811.

Parmley, K.A., Higgins R.H., Ganapathysubramanian B., Sarkar S., and Singh A.K. (2019) Machine Learning Approach for Prescriptive Plant Breeding. *Sci Rep* 9, 17132.

Scheffe, H. (1959) *The Analysis of Variance*. Wiley, New York.

Semenov, M.A. and Halford, N.G. (2009) Identifying target traits and molecular mechanisms for wheat breeding under a changing climate. *J Exp Bot* 60:2791–2804.

Tong, H. and Nikoloski, Z. (2021) Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data, *Journal of Plant Physiology* 257:153-354.

World Resources Institute - WRI (2021) Creating a Sustainable Food Future. Washington DC, USA.