HIGH THROUGHPUT PHENOTYPING AND MACHINE LEARNING FOR PLANT BREEDING

Karansher Singh Sandhu

Department of Crop and Soil Sciences, Washington State University, Pullman, WA, USA

The global human population is expected to increase by 25% over the next 30 years and reach 10 billion. For meeting the needs of food, fiber, and feed, new methods are needed for crop breeding to increase agriculture production. Since the last two decades, the field of genomics has revolutionized the field of plant breeding due to a reduction in genotyping cost, which results in the adoption of new technologies such as genome-wide association studies, genomic selection, rapid generation advance, gene editing, and speed breeding. For obtaining higher efficiencies from these technologies, accurate and reliable phenotyping methods are required. However, the field of plant phonemics has lagged with the advancement in the genomics. Phonemics refers to the study of plant growth, development, performance, and composition. The conventional field phenotyping used by the majority of plant breeders is labor-intensive, costly, subjective, and usually, a single measurement is taken at the end of the field season.

During the last decade, there is rapid adoption of the ground and aerial platforms with multiple sensors in agriculture for phenotyping various traits throughout the growth stage of the plant to alleviate these phenotyping bottlenecks. High throughput phenotyping (HTP) involves the application of tools for phenotyping the plants, which vary from the ground-based imaging to aerial phenotyping with unmanned aerial vehicles. These HTP uses technologies such as spectroscopy, non-invasive imaging, robotics, image analysis, and high-performance computing. All these tools vary in their capacity and cost but provide dense phenotypic data. HTP has opened the prospectus for non-destructive field phenotyping for the number of traits, including biotic, physiological, and abiotic stresses Both aerial and ground-based HTP are being rapidly adopted in agriculture in countries like the USA, Canada, and Australia for measuring multiple plant traits at different growth stages accurately and precisely.

Ground-based HTP platforms were developed for controlled environments and field applications such as robotic type and manually driven vehicle type. Most of the ground-based phenotyping platforms are used for indoor facilities like greenhouses and growth chambers with precise environmental control. Furthermore, several field-based HTP platforms such as tractor mounted systems, aerial vehicles, and pushcarts are recently adopted by plant scientists for phenotyping. These phenotyping platforms use a different number of sensors and height-adjustable sensors for measuring phenotypes over the various growth stages of the plants. Data collection from these sensors is stored in specific data loggers, general-purpose laptops to industrial computers.

The use of these HTP technologies has resulted in the problem of massive data analysis for the extraction of valuable information. Machine learning provides an alternative opportunity to the plant breeders, pathologists, and agronomists for the extraction of large number of parameters for analyzing each trait together, in spite of traditionally where we just used to look at a single feature at a time. Machine learning (ML) is an interdisciplinary approach for data analysis using probability, statistics, classification, regression, decision theory, data visualization, neural networks for relating information extracted with the phenotyped obtained

The other great breakthrough with ML is directly linking the variables extracted from the HTP data to the plant stresses, biomass accumulation, grain yield, and soil characteristics. The biggest success in ML involves inferring trends from the data and generalizing the resulted by training the model. The main driving forces behind the application of all these techniques in agriculture involves driving action by commercial companies and a reduction in the cost of sensors and imaging platforms

Typically, an ML model consists of a calibration process where a model is trained on a given large data set and is called a training set. The remaining dataset on which the model's performance is validated is called the testing set. The accuracy and precision of the model classify the use of the calibrated model for future applications. During the model training process, generally, two different supervised and unsupervised machine learning approaches are used. Supervised ML models involve where the label is provided for the data during the training process, for example, if you are differentiating wheat and rice with images, we provide the labels for these two crops while training the model using images. On the other hand, the unsupervised model does not involve the labels during the training process, meaning models try to differentiate both crops on its own by learning similarities and dissimilarities. There have been various ML models being applied for HTP, namely support vector machine, discriminant analysis, k means clustering, neural networks, clustering, dimensional reduction, and least discriminant analysis. All these models aid in the identification, classification, quantification, and prediction of different phenotyping components in plants.

It is concluded that application and scope of machine learning models for processing, extraction, and prediction of information during high throughput phenotyping. The use of these technologies is at the peak of the adoption by the research scientists for paving the way for food security.