

B-216

Machine learning approaches in soybean phenomics: predicting seed yield, oil and protein in contrasting production systems

Kyle Parmley^{*}, Department of Agronomy, Iowa State University, Iowa, USA

Race Higgins, Department of Agronomy, Iowa State University, Iowa, USA

Kevin Falk, Soybean Breeding and Genetics, Iowa State University, Iowa, USA

Johnathon Shook, Predictive Plant Phenomics Program, Iowa State University, Iowa, USA

Koushik Nagasubramanian, Department of Mechanical Engineering, Iowa State University, Iowa, USA

Baskar Ganapathysubramanian, Department of Mechanical Engineering, Iowa State University, Iowa, USA

Soumik Sarkar, Department of Mechanical Engineering, Iowa State University, Iowa, USA

Asheesh K. Singh, Department of Agronomy, Iowa State University, Iowa, USA

Genetic improvement of soybean [*Glycine max* (L.) Merr.] has permitted the expansion of soybean across a broad geographic region. Past breeding efforts have attempted to develop highly stable cultivars to deploy across all production systems, but these genotypes may evade an advantageous genotype by management (G x M) interaction, i.e., row width spacing. The development of these production system targeted cultivars will require continual improvement of yield per acre of soybean, which in turn will be dependent on the modification of physiological traits. Advances in remote sensing technologies have enabled rapid measurements of these traits on a temporal and spatial scale, and therefore are becoming increasingly adopted in advanced breeding systems. The objective of this study to develop yield prediction models using machine learning approaches. We used two independent studies with 32 genotypes of the SoyNAM panel with contrasting treatments: row width spacing (38 and 76 cm) and seeding density (123, 345, 568 x 10³seeds ha⁻¹) from nine environments in replicated tests. Physiological trait data of hyperspectral reflectance, leaf area index, canopy temperature, and chlorophyll content were collected at five time points during the growing season. Robust in-season prediction models identified informative explanatory variables for seed yield, oil and protein predictions, which will aide in breeding applications for contrasting production systems. Preliminary results indicate prediction accuracies were also similar for remote sensing tools with moderate and high throughput capability thereby decreasing the temporal requirement for data acquisition. The application of these approaches enable a mechanistic understanding of yield drivers in contrasting production systems and enable more informative decision making capability.