B-215

End-to-end phenotyping pipeline integrating computer vision and machine learning for genome-wide studies

*Jiaoping Zhang**, Department of Agronomy, Iowa State University, Iowa, USA
Hsiang Sing Naik, Department of Mechanical Engineering, Iowa State University, Iowa, USA
Alec Lofquist, Department of Computer Science, Iowa State University, Iowa, USA
Teshal Assefa, Department of Agronomy, Iowa State University, Iowa, USA

Accurate and efficient assessment of crop biotic and abiotic stresses is important for agricultural management practices and genetic studies. Traditional visual rating is time-consuming and labor-intensive, and lack of inter-rater reliability and intra-rater repeatability, which is further exacerbated by the necessity of multiple environments of replicated trials. Here we report a machine learning (ML)-enabled image-phenotyping pipeline that was used to evaluate the iron deficiency chlorosis (IDC) in soybean followed by genome-wide association study (GWAS). The pipeline includes field image capture data storage and curation trait extraction machine learning (ML)/classification models/apps for decision support. A set of over 4500 high quality canopy images representing different development stages of IDC symptoms acquired from a replicated trial consisting of 461 soybean germplasm lines was used to train ML models. A hierarchical classifier performed best among ten classifiers with a mean per-class accuracy of ~96%. We incorporated this workflow into a smartphone app that enables automated real-time evaluation of IDC scores using digital images of the canopy. The IDC severity was also calculated based on the automatically extracted IDC canopy trait features. The ML-generated IDC score (ML-score) and severity (ML-severity) was subsequently utilized for GWAS with 36,139 SNPs. The results illustrate the reliability and advantage of the ML-enabled phenotyping pipeline by identifying previously reported locus and a novel locus harboring a gene homolog involved in iron acquisition. This study provides a systematic framework that can be embedded onto ground vehicle and unmanned aerial system to allow high throughput phenotyping of stress-related traits during plant genetic research and breeding, and is an integral part of advancements in phenomics.