AG2PI SEED GRANT - PROJECT FINAL REPORT

PROJECT NAME	Machine Learning Competitions for G2P and End of Season Phenotype Prediction

PROJECT PRINCIPAL INVESTIGATOR	TODAY'S DATE	PROJECT START DATE	DATE OF COMPLETION
Abby Stylianou	8/14/2022	6/1/2021	6/1/2022
TEAM MEMBERS (co-PI, co-I, personnel)		COLLABORATORS	
Madison Pope (Graduate Student)			

ACCOMPLISHMENTS

We successfully ran two machine learnning competitions that relate to plant phenotyping and the genotype x phenotype relationship in crops and have organized a third to be run next year.

The two completed competitions were a Sorghum Biomass Estimation competition at the Workshop on Computer Vision Problems in Plant Phenotyping and Agriculture (CVPPA) at the 2021 International Conference on Computer Vision (ICCV) and a Sorghum Cultivar Classification competition at the Workshop on Fine Grained Visual Categorization (FGVC) at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). We have additionally planned a competition on predicting genetic markers from visual data that will be part of a planned submission to the Datasets and Benchmarks track at the 2023 Conference on Neural Information Processing Systems (NeurIPS) that has resulted from introductions made through this award.

The Biomass Estimation Competition received 9 separate submissions, with four teams performing notably above a simple predictor that merely predicted the average end of season biomass from the training data. The Sorghum Cultivar Competition received 252 separate submissions many of which achieved above-human performance on the recognition task.

We received useful feedback from contestants about their relative interest in the different tasks – machine learning practitioners were far more interested in the classification task as it fit existing pipelines and was more accessible, despite it being the less interesting task from a genotype x phenotype understanding perspective – as well as challenges with the data (and in particular the *scale* of the data – that large image datasets were a large barrier to entry) that would useful to understand for others intending to design machine learning competitions for genome and phenome understanding. While some contestants were excited to share their developed solutions, the prevelance of cheating and limited ongoing interest from competitors after their position on the public leaderboard has been secured present significant challenges in operationalizing machine learning competitions in improving scientific understanding.

Products

ACTIVITY / PRODUCT	DESCRIPTION (include URL, if applicable)	OUTCOME / METRICS
Sorghum Biomass Estimation Dataset and Competition – CVPPA 2021	https://www.kaggle.com/c/sorghum-biomass-prediction/leaderboard A publicly available dataset and competition consisting of nearly 300,000 overhead RGB images of biomass sorghum, annotated by plot-level biomass measurements. Competitors then predict the plot-level biomass for unlabeled data, with more credit given to submissions that make accurate predictions from earlier season data.	The Biomass Estimation Competition, hosted at the Workshop on Computer Vision Problems in Plant Phenotyping and Agriculture (CVPPA) at the 2021 International Conference on Computer Vision (ICCV), received 9 separate submissions, with four teams performing notably above a simple predictor that merely predicted the average end of season biomass from the training data. One of the top performing teams submitted and presented a write up of their approach to the workshop: "Pretraining on growth stage from images facilitates Sorghum Biomass prediction" by John A Fozard of the John Innes Centre.
Sorghum Cultivar Classification Dataset and Competition – FGVC 2022	https://www.kaggle.com/competitions/sorghum-id-fgvc-9/leaderboard A publicly available dataset and competition consisting of roughly 40,000 training overhead RGB images of biomass sorghum labeled by cultivar, and 40,000 testing images whose cultivar label was unknown. Competitors are tasked with predicting the cultivar for the unlabeled images.	The competition received 252 submissions, with 17 competitors achieving over 90% accuracy on the task (far exceeding human performance on the task).
What Does TERRA-REF's High Resolution, Multi Sensor Plant Sensing Public Domain Data Offer the Computer Vision Community?	https://terraref.org/sites/terraref.org/files/53.pdf A paper presented at the ICCV 2021 Workshop on Computer Vision Problems in Plant Phenotyping and Agriculture proposing a number of different tasks and competitions that would be of interest to the computer vision and machine learning communities using the TERRA-REF data (the dataset leveraged in the competitions run under this grant).	Recipient for the Best Paper Award (tied) at the ICCV 2021 Workshop on Computer Vision Problems in Plant Phenotyping and Agriculture. Additionally, as a result of this paper, we made a connection with the team at AICrowd (https://www.aicrowd.com/) and are planning to submit a genotype x phenotype benchmark and competitions to the Datasets and Benchmarks track at the 2023 Conference on Neural Information Processing Systems (NeurIPS).

Audience

This work has been targeted at relevant computer vision and machine learning communities through conferences and workshops that are focused on relevant research areas. Specifically, we ran competitions through the Workshop on Computer Vision Problems in Plant Phenotyping and Agriculture (CVPPA) at the 2021 International Conference on Computer Vision (ICCV) and the Workshop on Fine Grained Visual Categorization (FGVC) at the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). The competitions were publicly run on Kaggle, and the FGVC competition was promoted by Kaggle on their front page.

Challenges

The most significant challenge in this project has been achieving meaningful engagement with competitors in the machine learning competitions. While some contestants were excited to share their developed solutions, most competitors entirely disengaged after their position on the public leaderboard has been secured. Additionally, we dealt with fairly significant cheating in the competition with more engagement, and, based on the level of engagement seen in our two competitions, machine learning practitioners were more interested in the task that was directly suited to their existing pipelines (cultivar classification) than the more interesting, but also more challenging, task of phenotype prediction (biomass prediction). These factors present significant barriers for the operationalizing of machine learning competitions in achieving meaningful strides in genotype x phenotype understanding.

While we still believe there is merit in using machine learning competitions to improve such understanding, we believe significant effort would need to be undertaken to insure competitions are both scientifically meaningful and interesting to the machine learning community (as measured by engagement), to avoid cheating, and to, most importantly, maintain lines of communication with competitors after the competition ends in order to actually turn solutions produced within the competition into useful, generalizable approaches outside of the competition.