Cloud native pipeline automates analysis of plant hyperspectral image data

Background

- Image data analysis in biology is a complex task, and data processing is especially challenging for plant images.
- Hyperspectral imaging is an imaging modality that creates 3-dimensional data cubes representing the subject matter in its field of view. These hypercubes can be mined for information hidden in the cubes that may not be visible using other modalities.
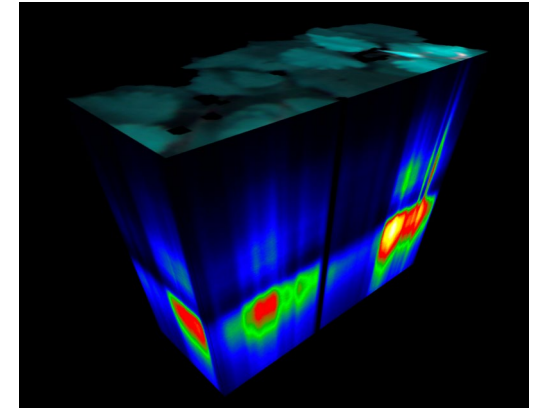
Approach

- We have developed a pipeline (HyperKube) that integrates machine learning and cloud native technologies to produce actionable information from multiple types of hyperspectral images of growing plants from ORNL's new Advanced Plant Phenotyping Laboratory (APPL).
- We used Onyx, an instance of Red Hat OpenShift's Kubernetes platform provided by ORNL's Office of Leadership Computing Facility (OLCF), to develop a pipeline for hyperspectral data ingest, extraction, and processing. Each element of the pipeline was encapsulated in a Kubernetes pod, which can be called on demand when new data arrive.
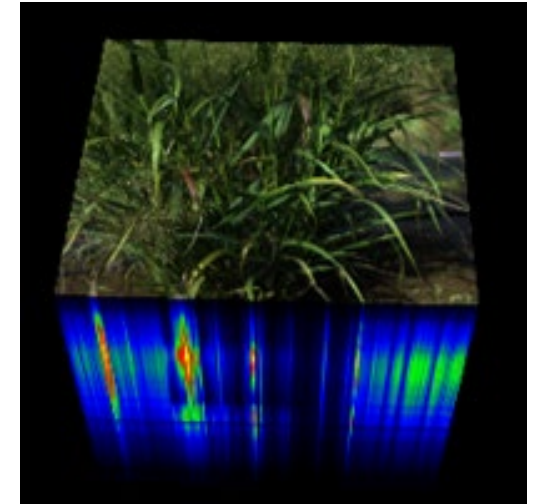
Results

- Hyperspectral data from APPL can now be automatically processed, with biologically relevant results delivered directly to principal investigators. The system has been expanded to all other APPL modalities so end users can compare results from different imaging stations. Now live, APPL users are already utilizing this pipeline in their experiments. The extendable method accommodates tailored analytical code written by other investigators.

Significance

- The data processing component of APPL, itself a highly automated image acquisition facility, has now been fully automated. This service enables users to obtain the full potential of what image analysis can achieve without becoming imaging specialists and accelerates investigators' productivity.

Yoon, H.-J. et al. ORNL/TM-2024/3279 (2024) https://doi.org/10.11578/dc.20231101.1, https://www.osti.gov/servlets/purl/2305378

A hypercube of *Populus trichocarpa* (poplar) leaves.

A hypercube of *Panicum virgatum* (switchgrass) whole plants.

U.S. DEPARTMENT OF ENERGY | Office of Science

THE CENTER FOR BIOENERGY INNOVATION

Biological and Environmental Research