The Ohio State University

ReRout Lab
Computer Systems Research That Reaches Out

Unlocking Planetary Digital Agriculture Transformation Using Intelligent CyberInfrastructure With Computational Learning in the Environment (ICICLE)

Christopher Stewart (Computer Science and Engineering),

Hari Subramoni (Computer Science and Engineering),

Scott Shearer (Food, Biological and Agriculture Engineering)

ICICLE AI Institute Thrust Leads on Use-Inspired Cyberinfrastructure for Digital Agriculture

Two Disciplines at Crossroads

Agriculture

Second agricultural revolution shifted from sustenance to business

- Tractors with large booms
- Chemicals for field management
- Labor issues

Third agricultural revolution introduced biological engineered of crops for pest resistance and improved yield

- Yield per hectare; efficiency
- Increase arable land

Population Growth to 9B by 2050
Cities growing; Arable land decreases

Computing

Early computing success in World War II, theoretical foundations; ENIAC

Early business uses: Mainframes, HPC, Programming languages, Moore's Law & Dennard Scaling, Minicomputers

Laptops, The Internet, Data Centers & Smartphones, Cloud Computing

- Democratization, data generation

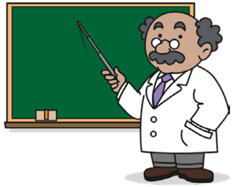
End of Moore's Law & Dennard Scaling
Saturated market for user attention

Digital Agriculture: The Hype

Technology Hype Curve

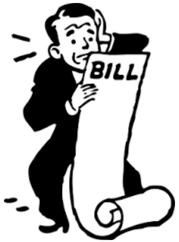
(adapted from Bill & Melinda Gates Foundation, ACM Communications 2021)

Innovation Trigger



Inflated Expectations

Translational Challenges

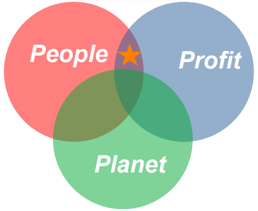


Early Products

HOPE

Impact & Adoption

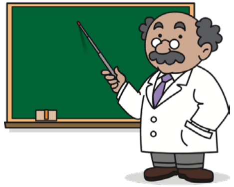
Sustainable Products



Digital Agriculture: The Reality

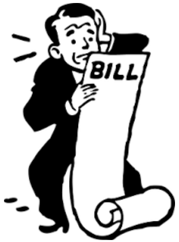
Only half of emerging technologies escape these stages

Innovation Trigger



Inflated Expectations

Translational Challenges

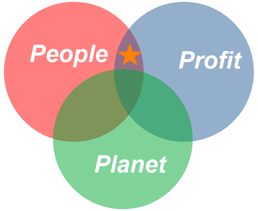


Early Products

HOPE

Impact & Adoption

Sustainable Products



Digital Agriculture: The Reality

Research & Innovation

arpa·e CHANGING WHAT'S POSSIBLE

+Novel world-changing tech
+Attract early-stage investors
- Path to impact is unclear

National Science Foundation

Translational Research

The Ohio State University
Translational Data Analytics Institute

+Convergent: Tackle big agricultural challenges
+Multiple disciplines combine to build technology and overcome obstacles
- Requires large teams & personal sacrifice

+Build user communities
+Develop trust for future tech
- Difficult to sustain innovation after early wins

Bill & Melinda Gates foundation

Early Deliverables & Immediate Impact

Digital Agriculture: The Reality

+Novel world-changing tech
+Attract early-stage investors
- Path to impact is unclear

+Convergent: Tackle big agricultural challenges
+Multiple disciplines combine to build technology and overcome obstacles
- *Requires large teams & personal sacrifice*

- Multi-disciplinary teams are expensive; market may not support solutions in long term
- Translation is slow, cumbersome, ad-hoc
- "Herding cats" phenomena

+Build user communities
+Develop trust for future tech
- Difficult to sustain innovation after early wins

Research & Innovation

Translational Research

Early Deliverables & Immediate Impact

arpa-e CHANGING WHAT'S POSSIBLE

National Science Foundation

The Ohio State University Translational Data Analytics Institute

Bill & Melinda Gates foundation

Digital Agriculture: Our Approach

- Multi-disciplinary teams are expensive; market may not support solutions in long term
- Translation is slow, cumbersome, ad-hoc
- "Herding cats" phenomena

Our Research Agenda for Digital Agriculture Systems

- **Data Driven**
  - Operate without human intervention
  - Resilient and reliable execution in diverse contexts
- **Democratized**
  - Highly accessible to technophiles and tech-agnostic
  - Trusted & Easy to use; humans teaming up with machines
  - Extremely affordable: Delivered as a free service

ICICLE-Centric Digital Agriculture

Research Challenge: *Be cost effective*

- *HW affects affordability*
- *Sensing and processing data using AI is computationally intensive*

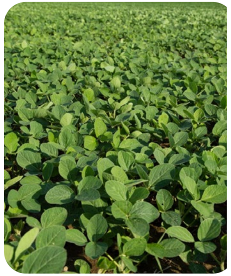
Research Challenge: Ag-Aware Design

- *Trusted systems must consider the application domain*
- *Domain specific systems are harder to use, less affordable*

Research Challenge: *Evaluate in the Wild*

- *Resilience and reliability benefit from realistic tests*
- *Accessibility and affordability suffer under long development cycles*

ICICLE Digital Agriculture High-Level Overview



-- How can we bridge the gap? --

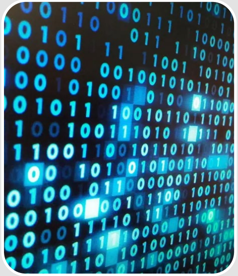
Crop Health Diagnosis

Crop Management Interventions

ICICLE Digital Agriculture High-Level Overview

Use-Inspired Cyberinfrastructure

Crop Health Diagnosis



HPC-Enabled Data Harvesting & Preparation

HPC for Ag AI

From HPC to the field: Computing Continuum

Crop Management Interventions

End-to-end software workflows that democratize access, use, and benefits of AI-driven digital agriculture

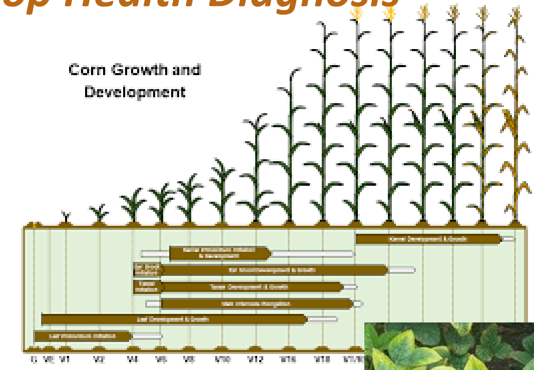
ICICLE Workflow for Digital Agriculture

Crop Health Diagnosis

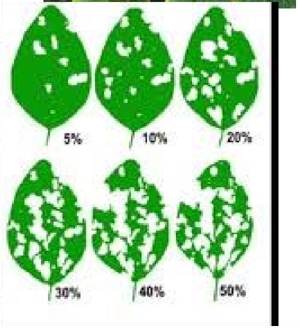
Partnerships
farmers, co-ops, institutes, etc.

Issues for Crop Health

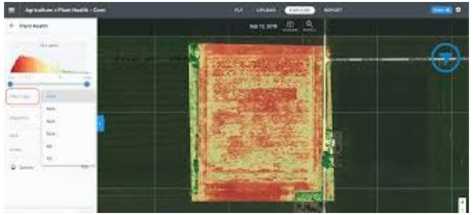
- Residue Cover on Soil Surface
- Soil Aggregate Size
- Crop Development
- Non-Uniform Emergence
- Nitrogen Deficiency
- Defoliation
- Crop Blight



a) Crop growth stages
b) Nitrogen deficiency
c) Crop expert Kelley Tilmon



Crop Management



a) Drone Deploy's Crop Health Mapping System
b) John Deere's See & Spray Selective Sprayer

Corrective Action

- Selective Spraying
- Nitrogen Use Efficiency
- Yield Estimation
- Replant Decisions
- Tillage Control

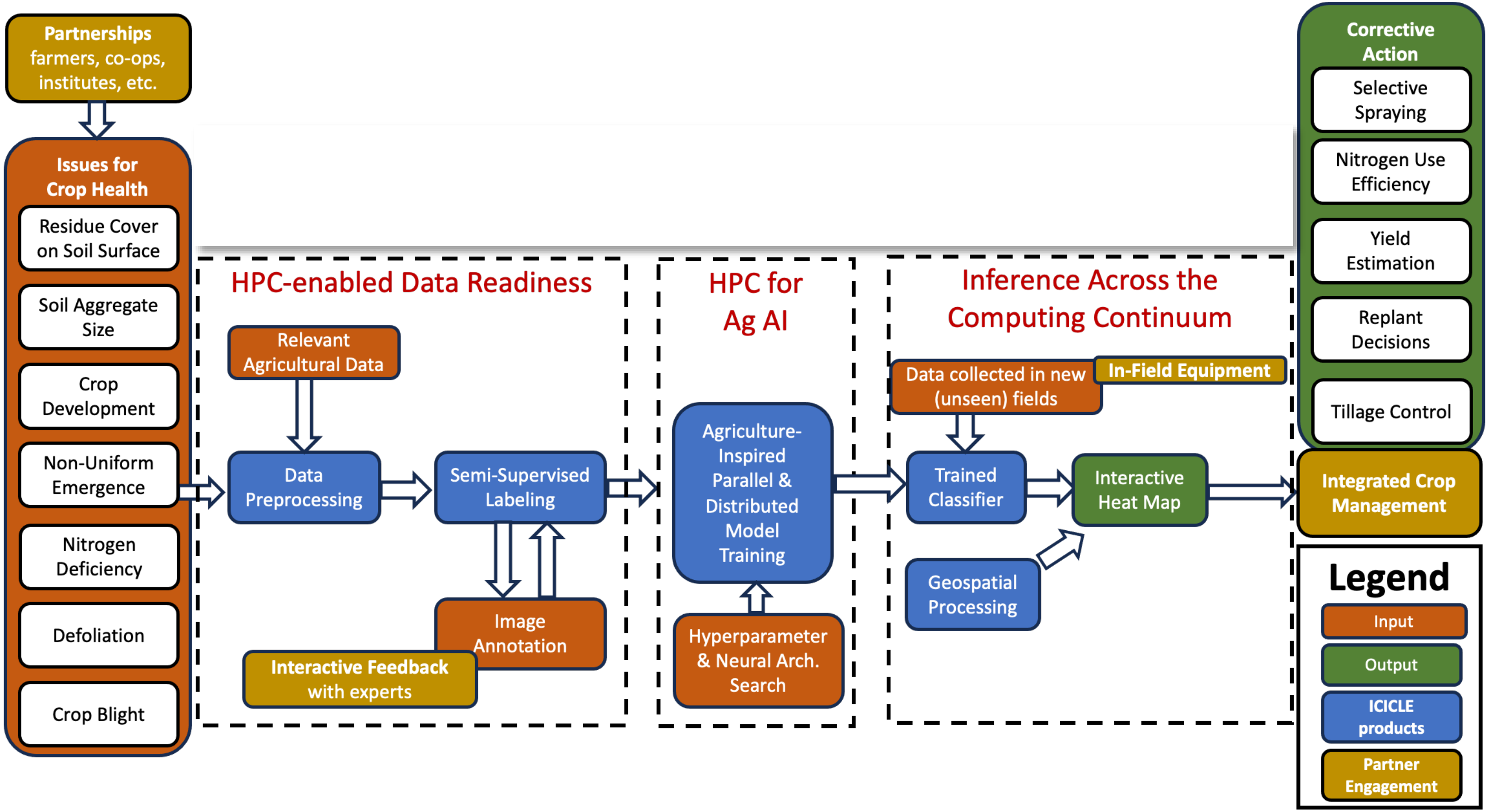
Integrated Crop Management

Legend

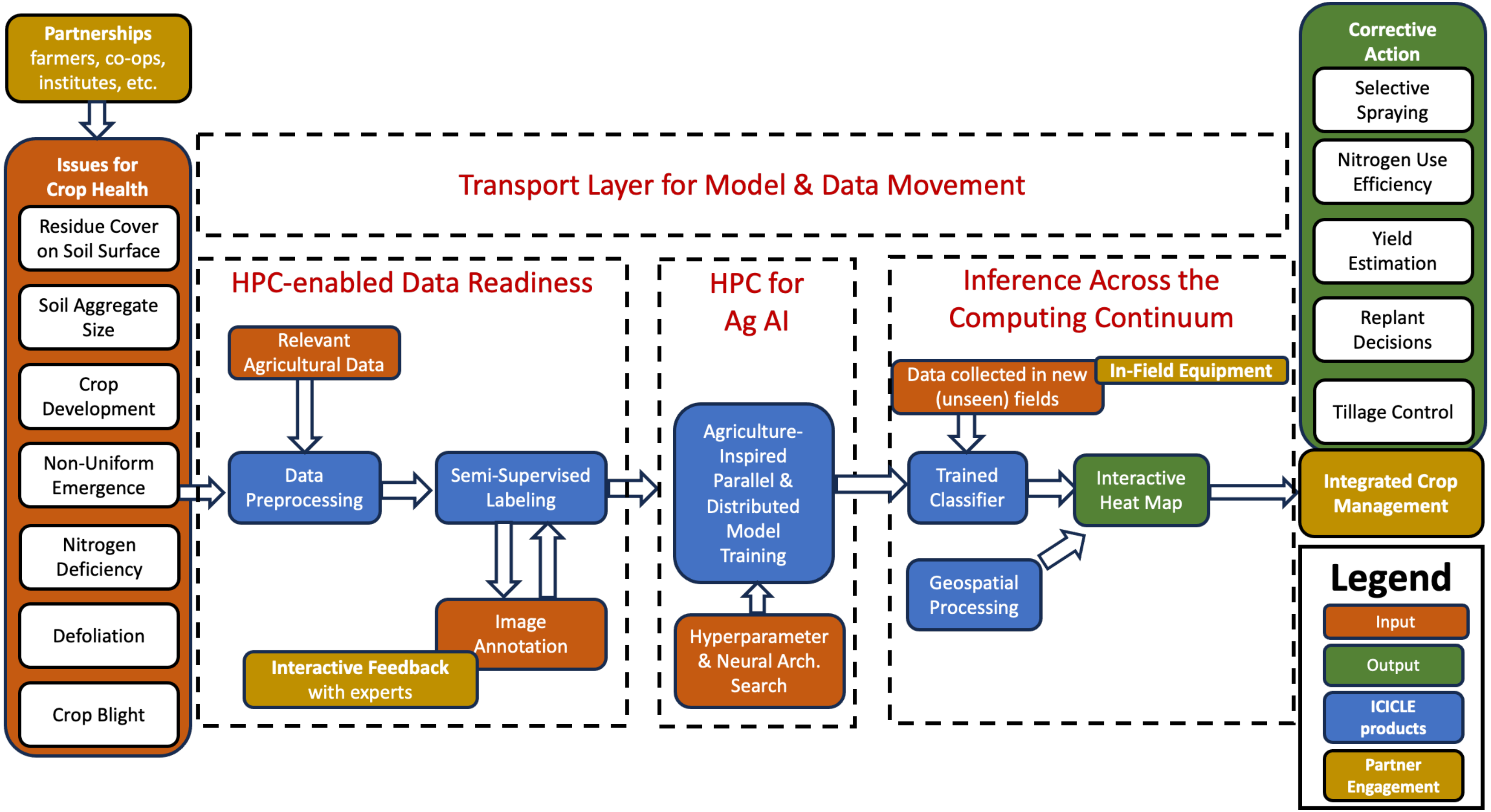
- Input
- Output
- ICICLE products
- Partner Engagement

ICICLE == DEMOCRATIZATION

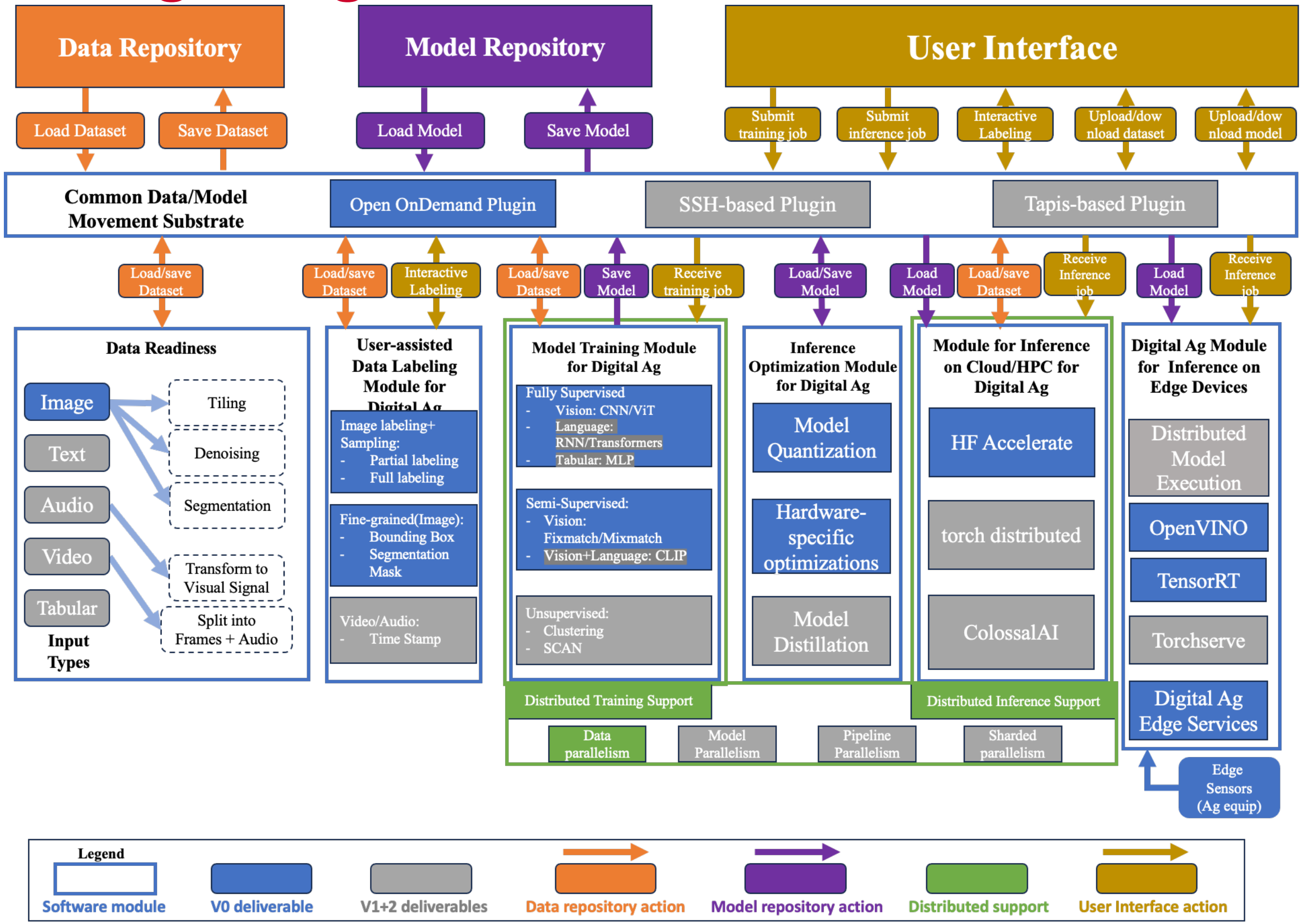
ICICLE Workflow for Digital Agriculture



ICICLE Workflow for Digital Agriculture



ICICLE Digital Agriculture Reference Architecture



Collaborating with ICICLE on Digital Agriculture

ICICLE + Partnerships = Democratization

- **The product you are developing uses the same machine learning technique that I use. Are we competitors? If so, why should I collaborate?**

- **ICICLE is a partner. Our focus is novel cyberinfrastructure. We want to share our work with the community. We also have capabilities to work confidentially with partners.**
  - **Let's grow the pie by making digital agriculture technologies (including yours) more accessible than they would be without intentional CI design and development**

Partners + ICICLE Reference Architecture = Use Cases

- During this talk, you showed some really cool products. I could have my students/staff replicate the work or I could use your tools. How do I access them? Can you customize them?

- Throughout the presentation, we will refer to project leads associated with each product. You can email:
  - Project leads directly
  - Scott Shearer, Chris Stewart, and Hari Subramoni (Digital Agriculture leads)
  - Neelima Savardekar and DK Panda
- Customizing products to your use case. Let's do it!
  - As special partners, our sister AI institutes can send representatives to our digital agriculture meetings to inform design and development directions

Partners + Data = Impact

- I see that you have access to and are informed by real digital agriculture data. That's great. But I have DATA (pronounced day-ee-tuh) and use cases that don't exactly match your services.

- Recall ICICLE seeks to build novel CI for a wide range of digital agriculture use cases. Novel and large datasets are welcome. We will help drive such projects directly:
  - Optimize machine learning pipelines (Hari)
  - Develop end-to-end workflows from diagnosis to intervention (Chris + Scott)
  - Optimize data management and data readiness workloads (Raghu)
  - Efficiently manage data movement across the cloud continuum (Spyros)
  - Contact: Scott Shearer, Chris Stewart, or Hari Subramoni (Digital Agriculture leads)
    - We will meet one-on-one and devise a plan to integrate your data

Data Driven + Use Inspired = Partner Driven

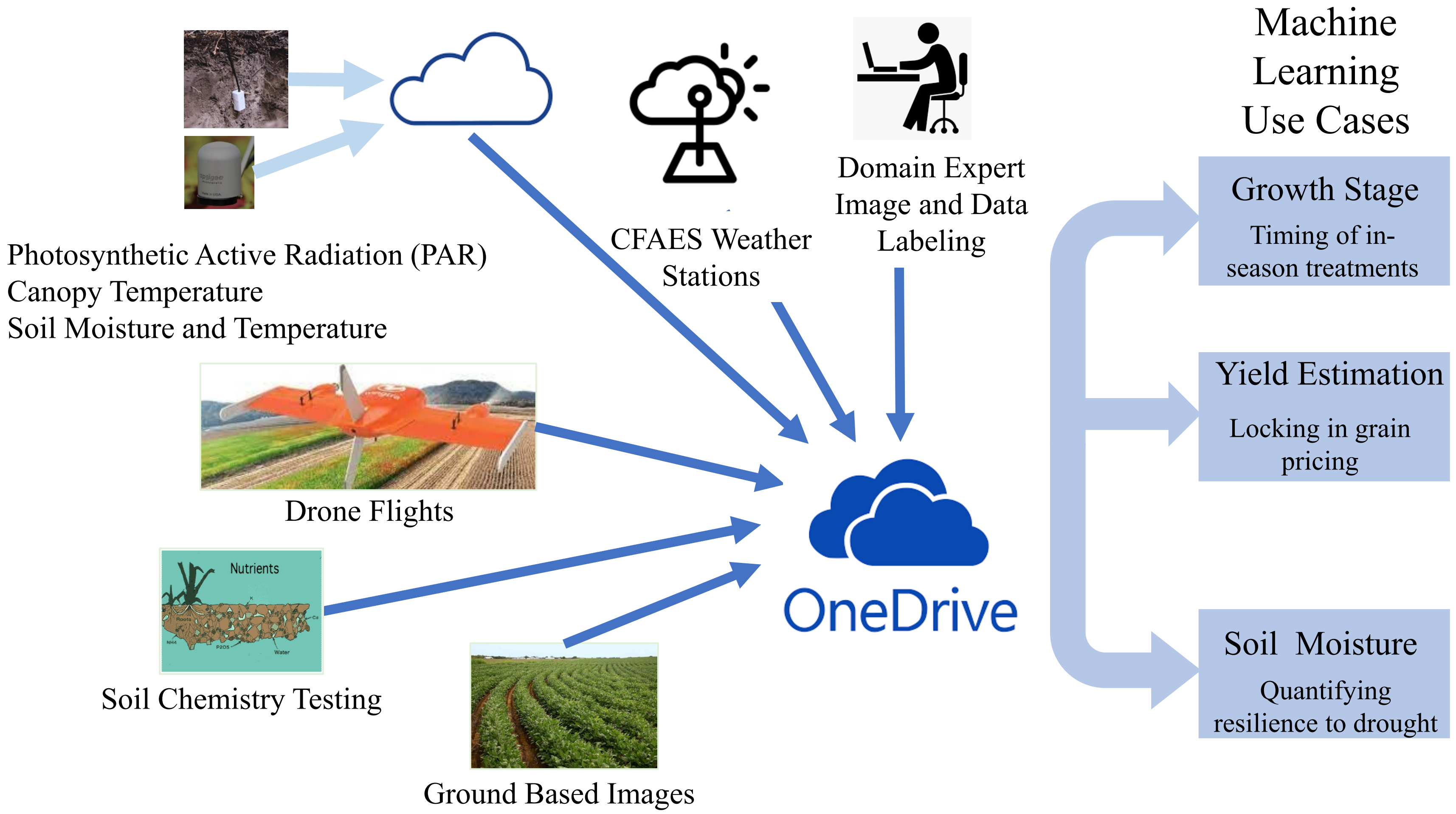
- Great services. Have you considered....

- Design insight: Our service-oriented architecture allows for AGILE software development practices. When services are released, human resources are freed. We are constantly re-prioritizing products and resources to build novel CI for a wide range of digital agriculture uses.

- Come to a meeting, share your idea and let's discuss integration.
  - Contact: Scott Shearer, Chris Stewart, or Hari Subramoni (Digital Agriculture leads)

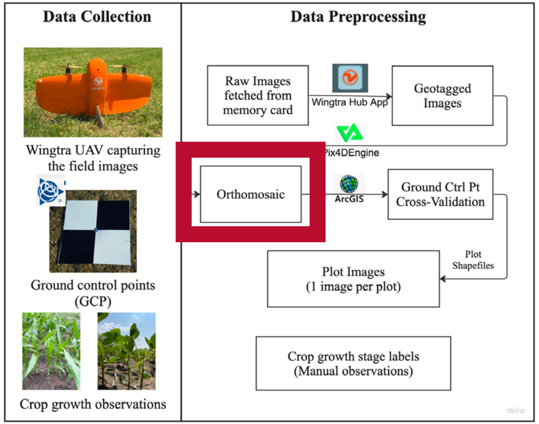
HPC-Enabled Data Readiness

Data Sources and Use Cases



CFAES

Orthomosaic Creation



<u>Orthomosaic creation is the biggest pain point in sUAS processing</u>

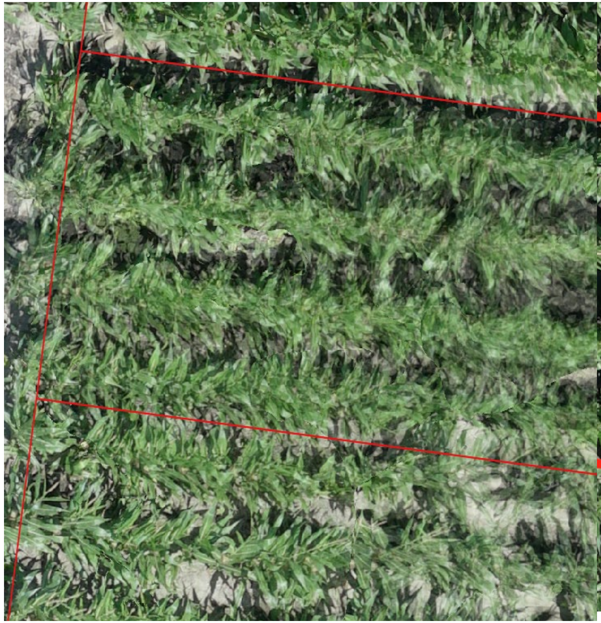
Best case

1. Upload 10-20GB of images (~6 hours)
2. Run Pix4D Engine script (~4 hours)

Reality – 27 of 85 flights from 2023 still haven't been processed

CFAES

Orthomosaic Quality

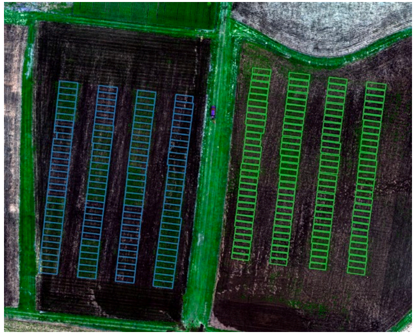


Orthomosaic image

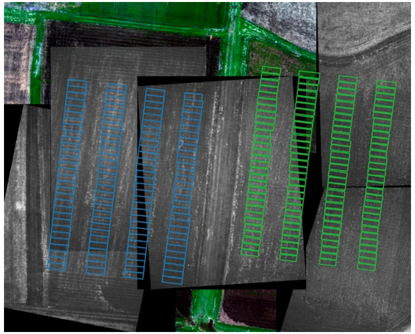
Original image

Our Approach

- Use an approach called **<u>direct georeferencing</u>** based on sUAS orientation at time of image capture to convert a PPK geotagged image directly to a georeferenced image (GeoTIFF) **<u>quickly without loss of image quality</u>**. **<u>(Goal of minutes instead of days)</u>**
- Direct georeferencing has **<u>up to .8m error</u>** versus typical orthomosaic accuracy of **<u>2-6cm</u>**.
- Enables reduced HPC processing time for training and allows for edge inference.



Existing orthomosaic image



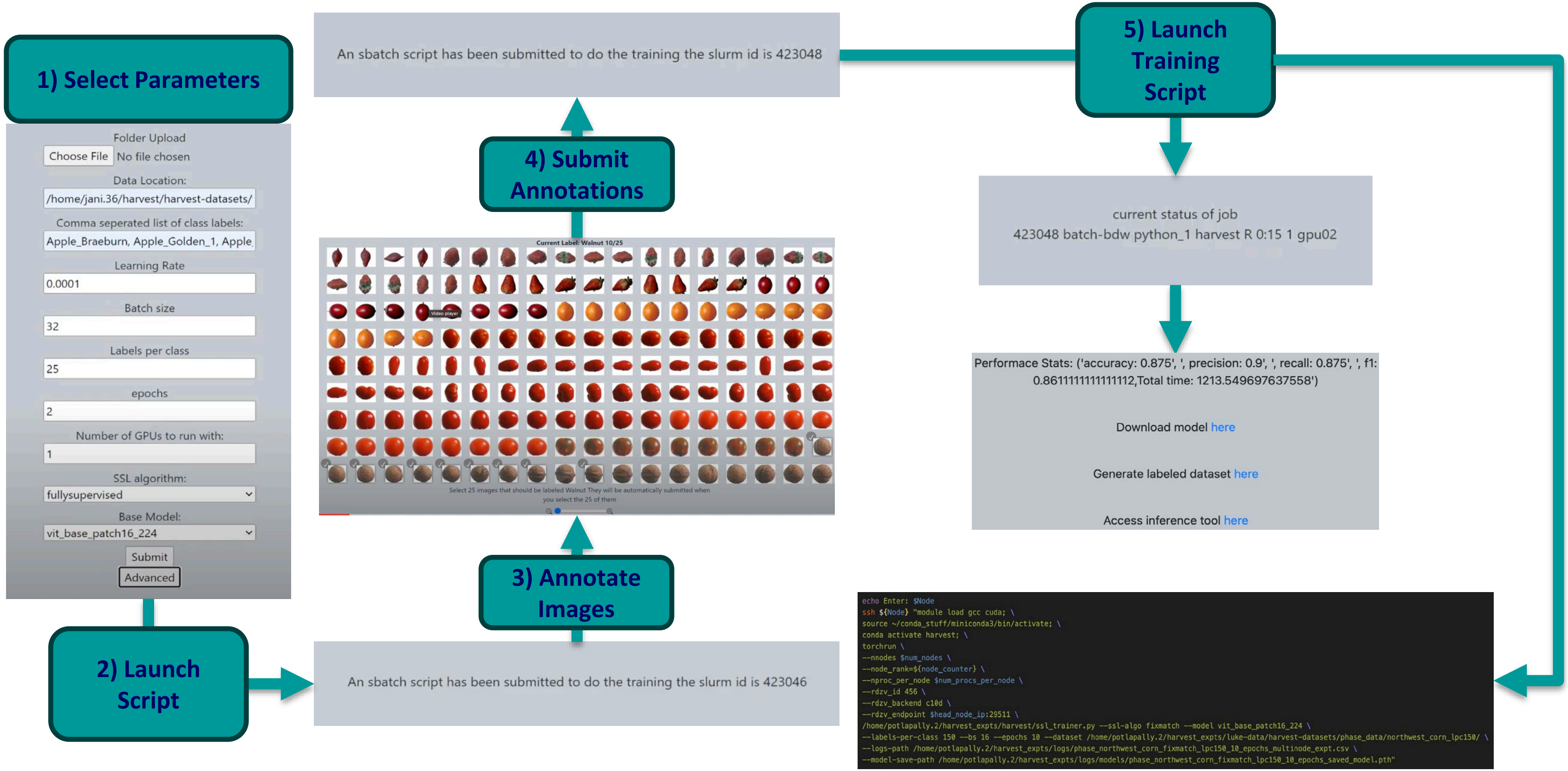
GeoTIFF images

Winter Extension Meeting 3/6 in Newark, OH

Feedback from farmers

- General negativity towards expensive technologies
- However, a lot of interest in “better managing data”, “using AI based insights”.
- Very interested in in-season yield estimation

Distributed Semi-supervised Learning Pipeline



Demo: Interactive Labeler

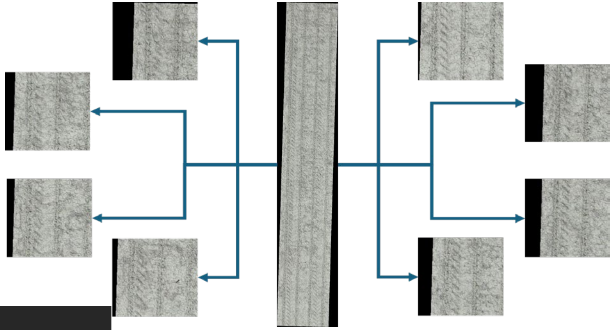
HPC for AG AI

Example Use Case: Crop Growth Stage Classification

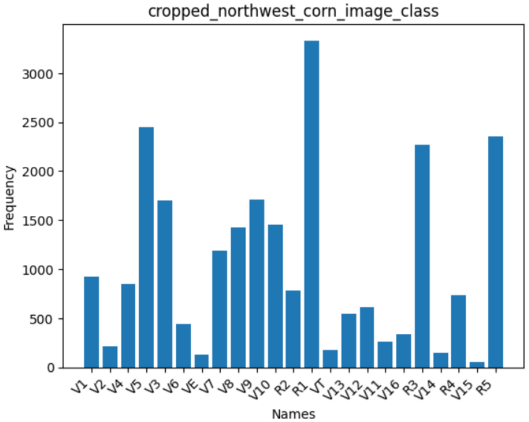
Goal: Train a classifier to identifies growth stages of crops in a field.

- **Data Pre-processing**
  - Obtaining image crops of size 224x224 for downstream training process
  - Exclude regions with blank pixels ([0,0,0]) during extraction process.
  - Generated dataset consists of 26,153 non-overlapping image crops

- **Directory Restructuring:**
  - Given a dictionary of images and a list of labels, separate into labelled and unlabeled images
  - Generate training and validation sets from labeled subset
  - Automate the process of data directory restructuring for model training



Training Scheme #1: Individual Growth Stages

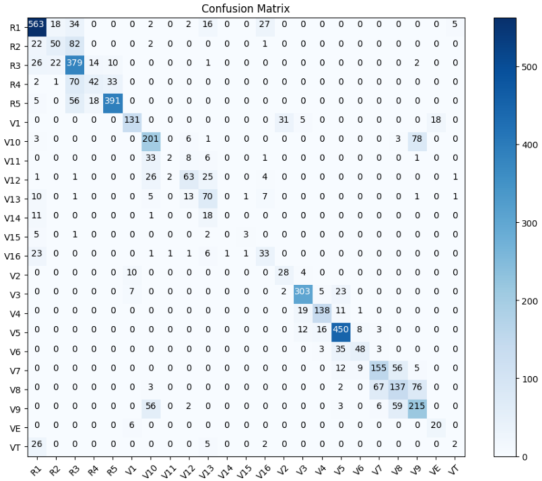


Class distribution of 23 growth stages of corn

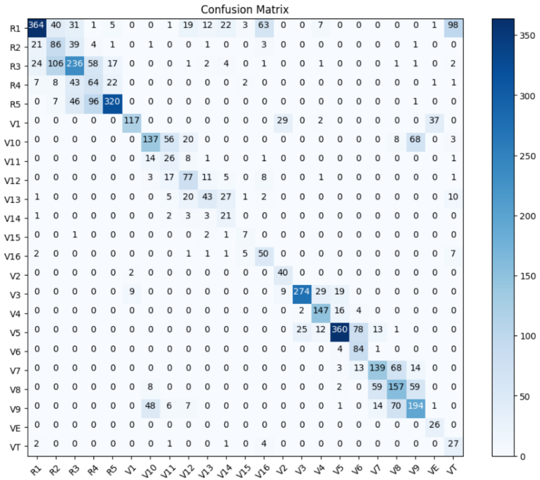
- **Trained using Vision Transformer (input size 224, patch size 16) with 23 classes both fully supervised and SSL.**
- **Reasonable accuracy for both fully supervised and SSL training given the number of classes.**
- **However, there's significant data imbalance.**
- **Furthermore, there's high visual similarity between temporally adjacent growth stages.**

Training method	accuracy	precision	recall	f1-score
Fully Supervised	0.71	0.6	0.56	0.56
FixMatch (#labels=50 per class)	0.62	0.55	0.66	0.57

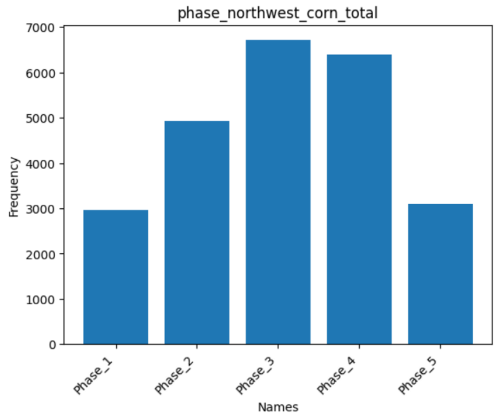
Fully Supervised



Semi-Supervised



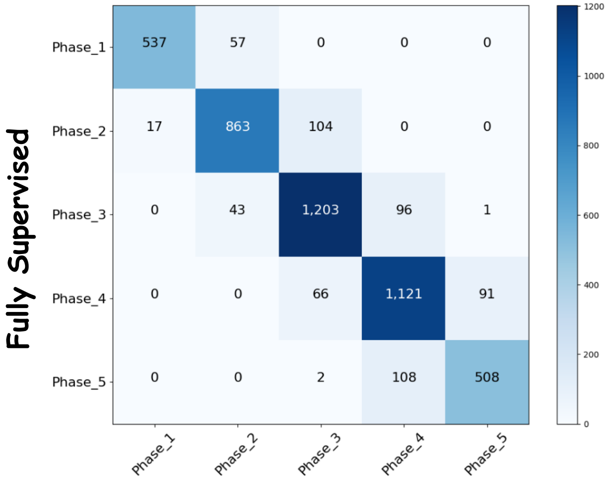
Training Scheme #2: Bucketized Growth Stages

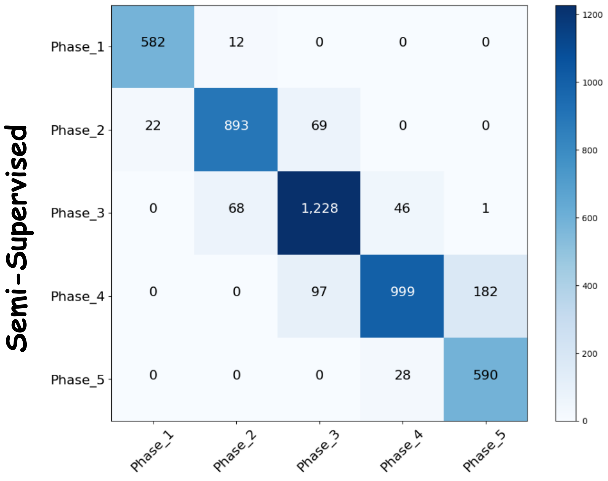


Class distribution of 5 bucketized growth phases

Training method	accuracy	precision	recall	f1-score
Fully Supervised	0.88	0.89	0.88	0.88
FixMatch (#labels=150 per class)	0.89	0.89	0.91	0.9

- **Growth stages can be classified into 5 phases, so that visually similar growth stages are bucketized together.**
- **This reduces the data imbalance.**
- **It also improves the overall accuracy at the cost of more coarse-grained classification.**





Agriculture-Inspired AI

Is this a soybean healthy field?

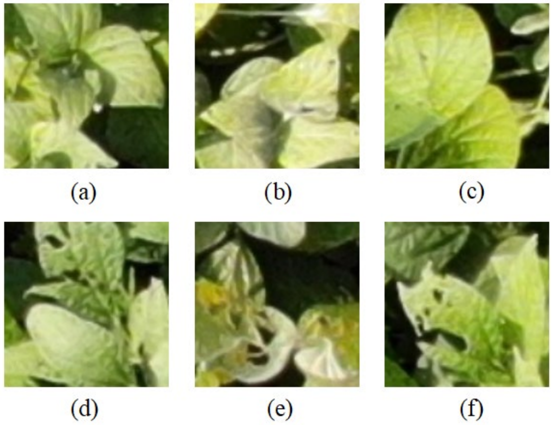
Image from Molly Caren Agricultural Center, OSU's 2100 -acre facility

Zichen Zhang, Sami Khanal, Amy Raudenbush, Kelley Tilmon, Christopher Stewart, *Assessing the efficacy of machine learning techniques to characterize soybean defoliation from unmanned aerial vehicles*, Computers and Electronics in Agriculture, 2022

Funded, in part, by the Ohio Soybean Council

Agriculture-Inspired AI

- Sub-problem: Are these healthy soybean?
- Soybean defoliation is correlated with insects and poor yield.
- *Can we use AI to detect defoliation?*



Agriculture-Inspired AI

a
b
c

1 2 3

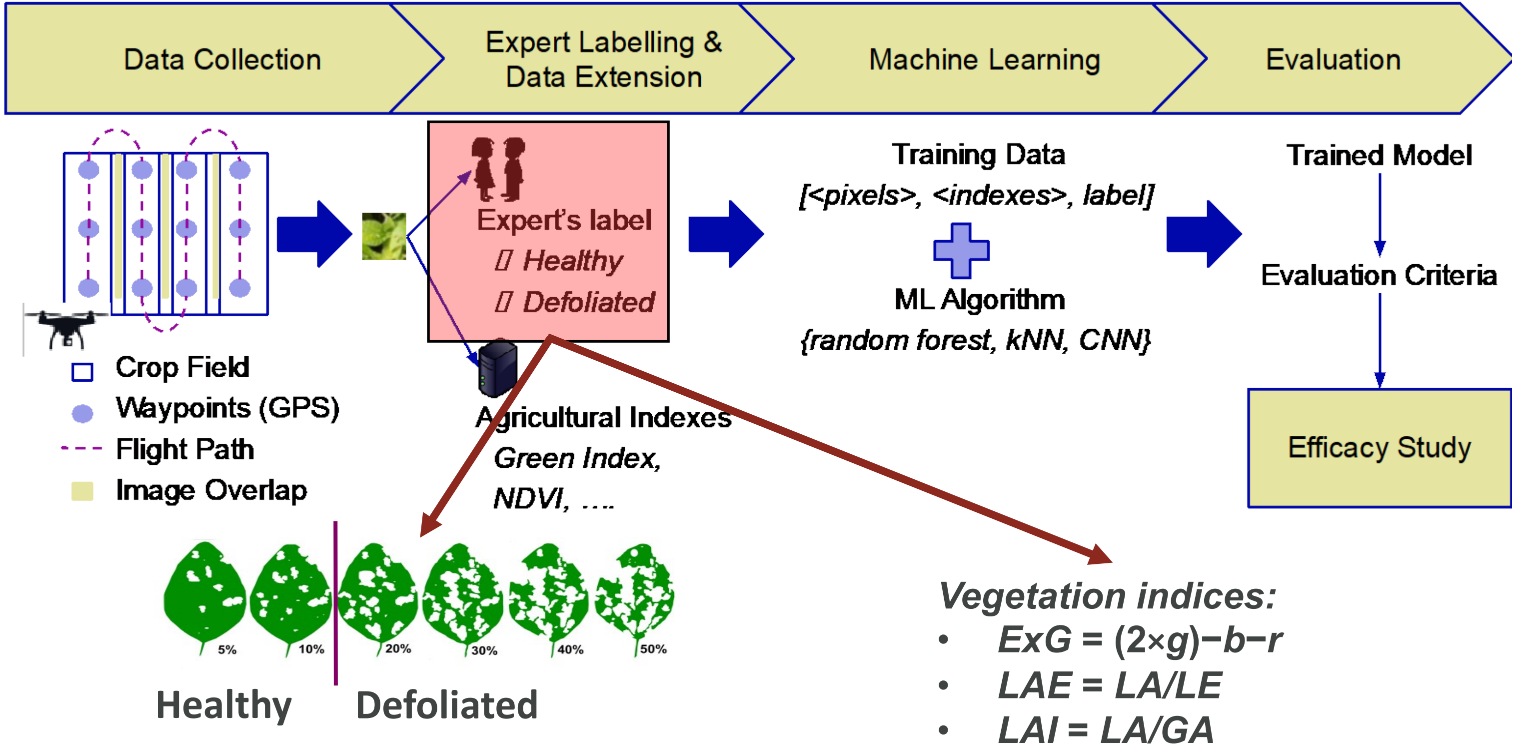
- Even with high-res images, classification is hard
  - Multiple leaves in an image
  - Overlapping leaves
  - Shadows/Ground

Should we develop AI models that are tailored to agricultural contexts or use off-the-shelf tools?

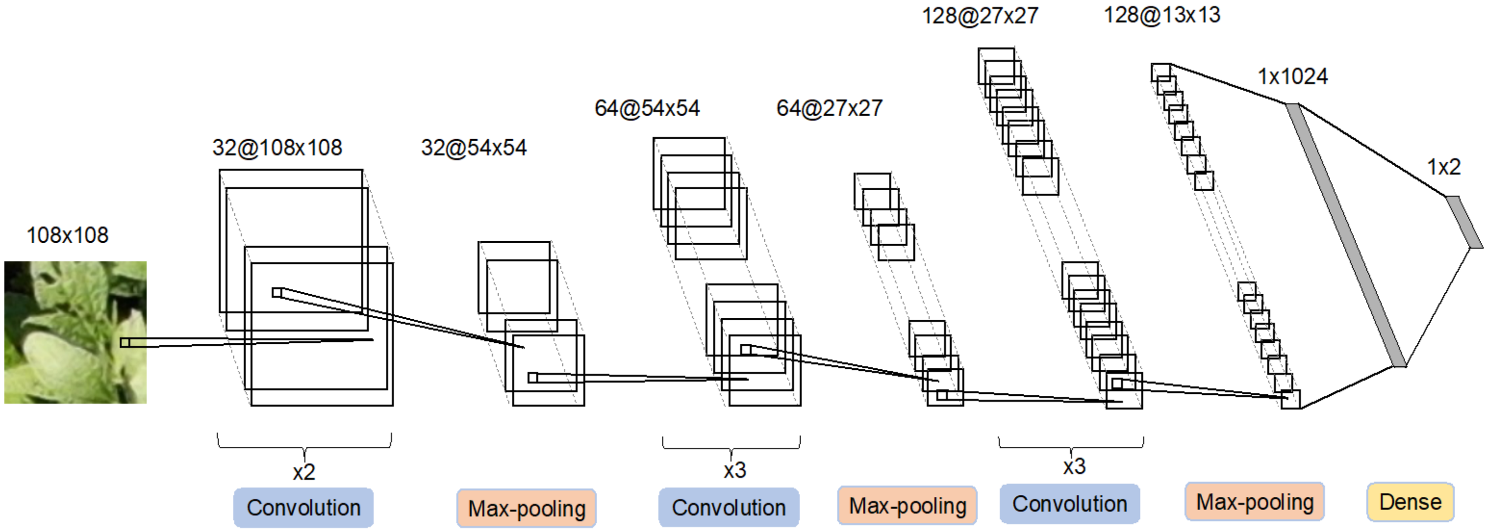
Are their specific usage scenarios in agriculture to consider?

Agriculture-Inspired AI

4-stage process for a study on the efficacy of machine learning for crop defoliation.



We built and trained DefoNet, a neural network architecture, to label severe defoliation in high-res aerial images



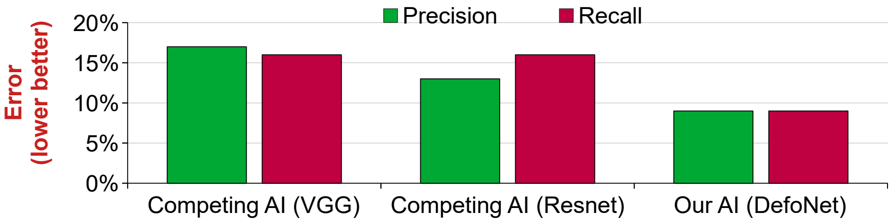
Agriculture-Inspired AI

I need to spray areas with severe defoliation

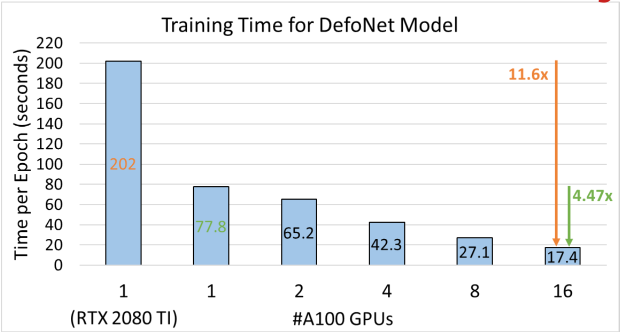
Train for Precision means label uncertain areas as defoliated

I want to see the portions of my field that are defoliated

Train for Recall means focus on areas labeled with high confidence

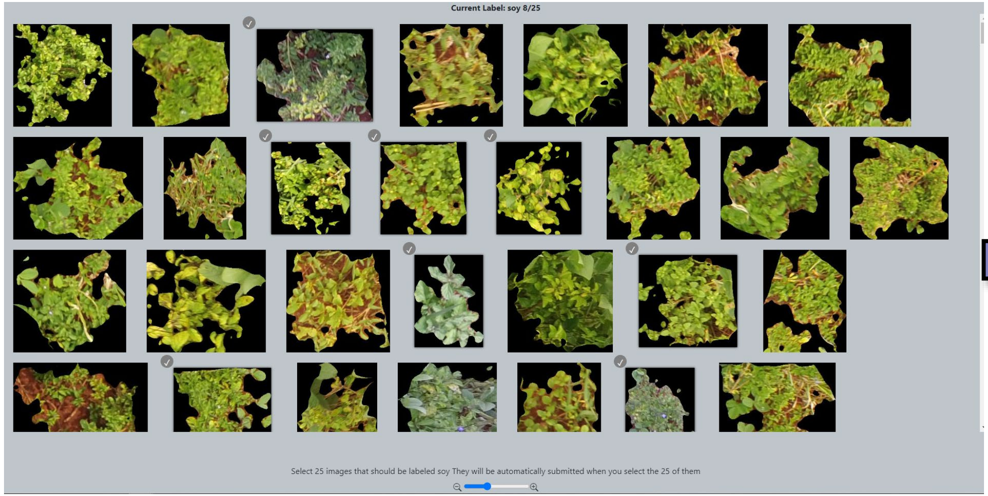


Can be trained affordably.



Recall with Hi-Res UAV	**91%**
Recall with Lo-Res	**73%**

Automatic Job Submitter



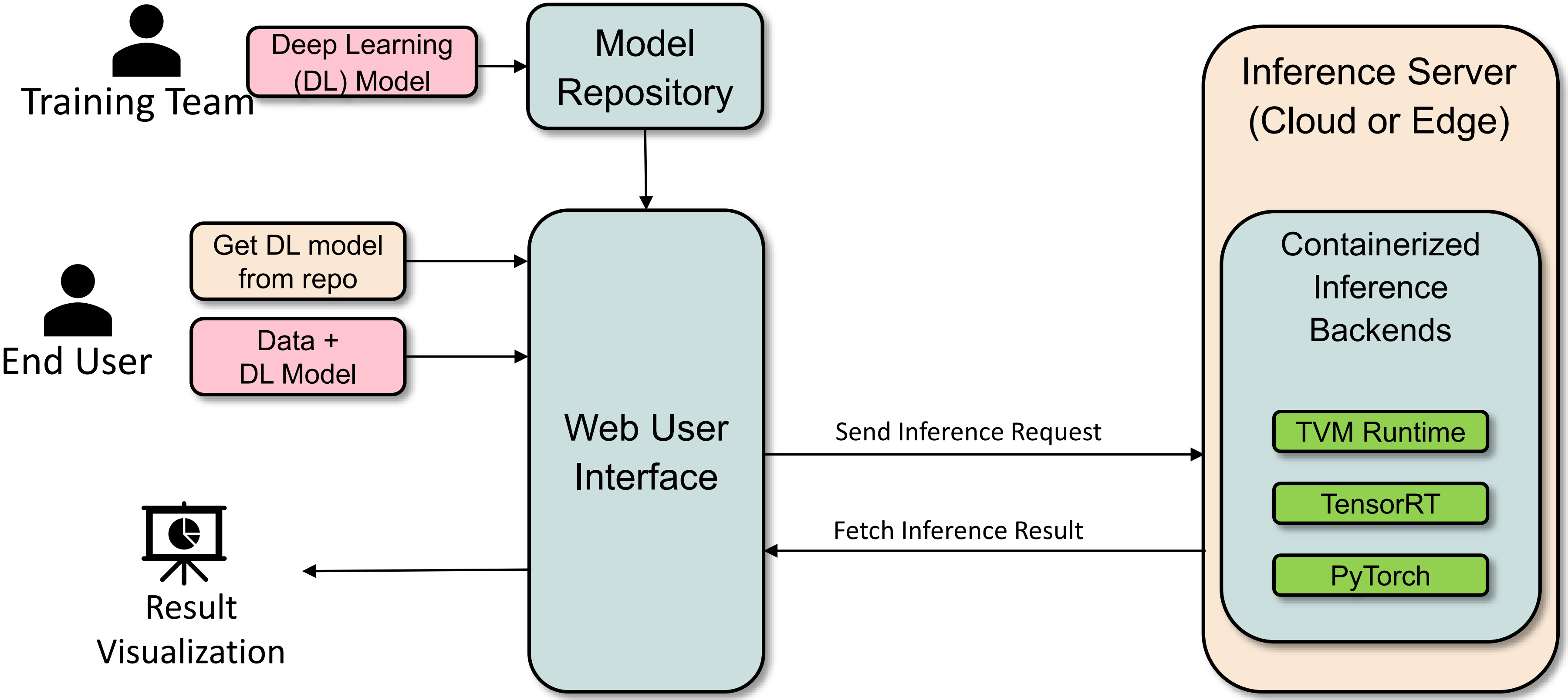
current status of job

423063 bdw-v100 python_1 harvest R 0:24 1 gpu12

The Job submitter takes information gained from the user and creates a script that will be submitted to the HPC resource for training or inference. While the job is in queue or running the user can see the status of the job as shown in picture above through the UI

Inference Across the Computing Continuum

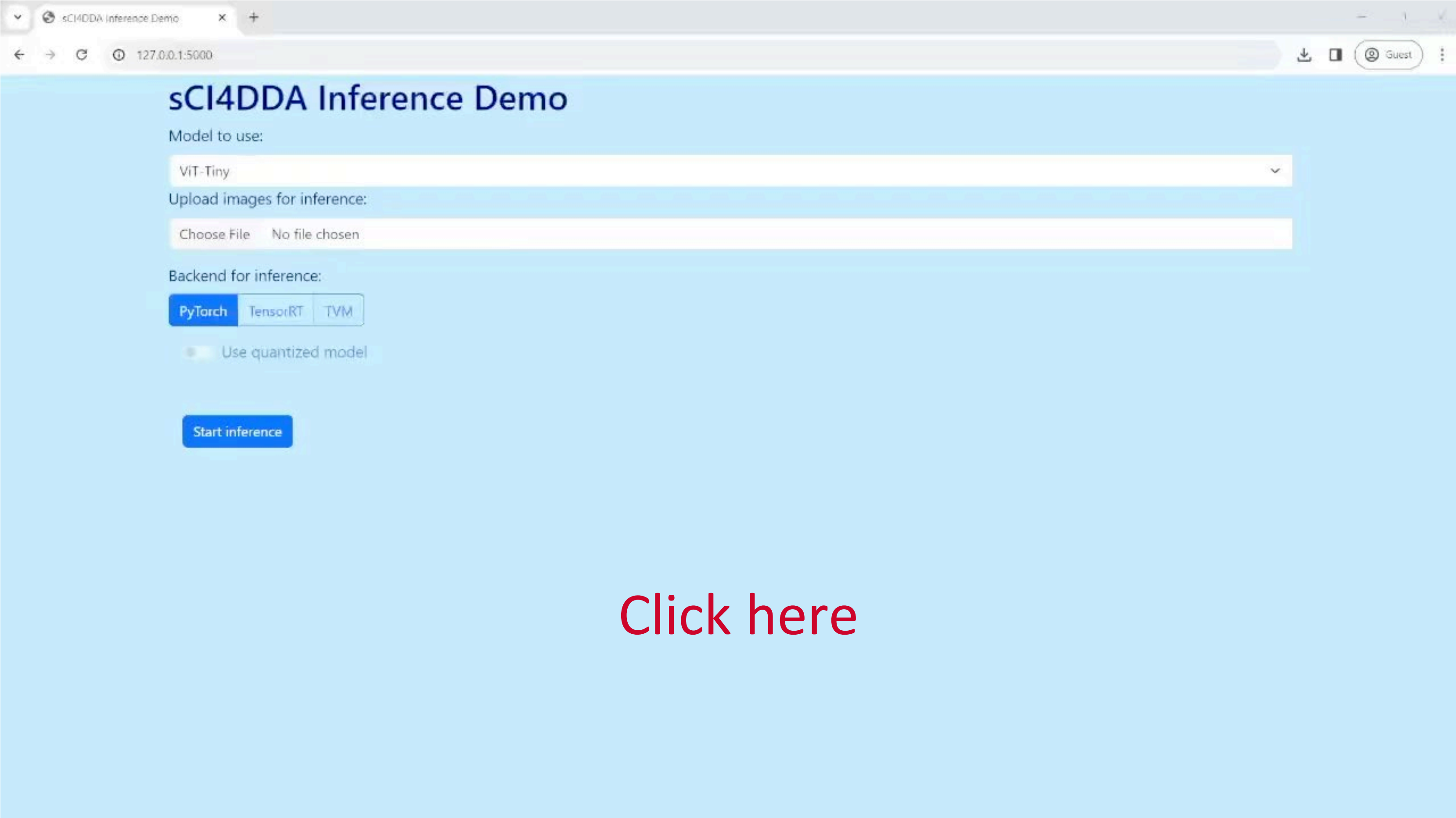
Accomplishments: Inference Across the Computing Continuum – Workflow



Accomplishments: Inference Across the Computing Continuum – Updates

1. **Containerized Inference Environment Deployment:**
   - Successfully deployed containerized inference environments on edge computers and ICICLE servers.
2. **Web-Based Inference UI Development:**
   - Created a user-friendly web interface allowing end users to upload datasets, models, execute inferences, and visualize results.
3. **Vision Model Inference Workflow:**
   - Engineered a specialized workflow supporting vision models like ViT and ResNet.
4. **Inference Backend Support:**
   - Implemented compatibility with PyTorch, TensorRT, and TVM runtime for diverse inference backends.
   - Initial support added for quantized model inference via TensorRT.
5. **Inference Throughput Profiling:**
   - Introduced preliminary support for profiling inference throughput, laying the groundwork for Q2 characterization analysis.

Accomplishments: Inference Demo

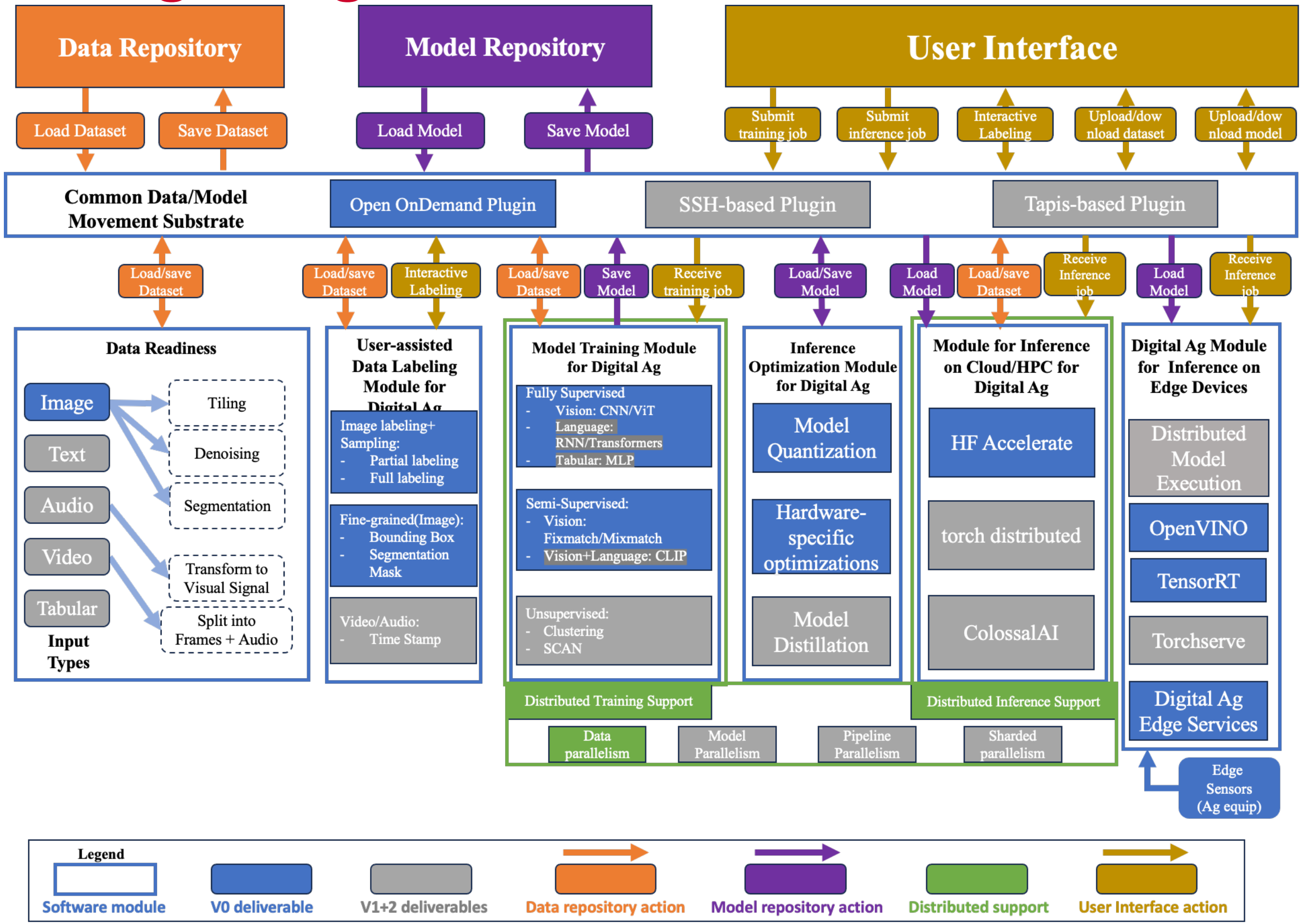


Click here

Demo: End User Visualization

Click here

ICICLE Digital Agriculture Reference Architecture



ICICLE Digital Agriculture – List of Technologies

Data Repository		Model Repository		User Interface
Data Readiness Module	Data Movement Substrate	User assisted data labelling and training modules	Inference on Edge and cloud/ HPC module	Stakeholder Engagement User Interface
wingtra	HDF5	PyTorch	docker	Flask
PIX4D	tapis	Semilearn	PyTorch	Jinja
ArcGIS	aws	Pydeck	TensorRT	slurm
rasterio	Google Cloud	GeoPandas		OPEN OnDemand
	Azure			JS
	openstack			

Team Members

- Overall Leads
  - Chris Stewart
  - Hari Subramoni
  - Scott Shearer
- Area Leads
  - Data Readiness
    - Raghu Machiraju and Luke Waltz
  - Distributed Model Training
    - Mustafa Abduljabbar
  - Inference Across the Computing Continuum
    - Aamir Shafi
  - Data Movement Substrate
    - Spyros Blanas
  - Stakeholder Engagement
    - Matt Lieber
  - BIBN
    - Chris Stewart
  - Reference Architecture
    - Aamir Shafi

Team Members

- Overall Leads
  - Chris Stewart
  - Hari Subramoni
  - Scott Shearer
- Area Leads
  - Data Readiness
    - Raghu Machiraju and Luke Waltz
  - Distributed Model Training
    - Mustafa Abduljabbar
  - Inference Across the Computing Continuum
    - Aamir Shafi
  - Data Movement Substrate
    - Spyros Blanas
  - Stakeholder Engagement
    - Matt Lieber
  - BIBN
    - Chris Stewart
  - Reference Architecture
    - Aamir Shafi

Interested in what we can build for you?

Join our meetings via Zoom.

Provide direct feedback on development plans, products, priorities

Contact: stewart.962@osu.edu or subramoni.1@osu.edu