Overview of the Digital Agriculture Activities in the ICICLE NSF-AI Institute

(ICICLE: Intelligent CyberInfrastructure with Computational Learning in the Environment)

Digital Agriculture Workshop, held in conjunction with HiPC ‘23

by

Dhabaleswar K. (DK) Panda
The Ohio State University
E-mail: panda@cse.ohio-state.edu
http://www.cse.ohio-state.edu/~panda

http://icicle.ai
Outline

- ICICLE Vision and Goals
- Research Challenges being Addressed
- Selected Accomplishments in the Digital Agriculture
- How to Get Engaged?
Computing has been evolving over the last three decades with multiple phases:

• Phase 1 (1975-): Scientific Computing/HPC
• Phase 2 (2000-): HPC + Big Data Analytics
• Phase 3: (2010-): HPC + AI (Machine Learning/Deep Learning)
Emergence of the Computing Continuum

1. Scientific Computing
2. Big Data & Data Science
3. Artificial Intelligence

HPC Systems & Data Centers

On Field Sensors
Edge & Near Edge
Hybrid Cloud

DA-Workshop (HiPC '23)
AI-Driven Digital Agriculture

https://ccag.tamu.edu/research-project/digital-agriculture/

Broad Challenge

Designing the next-generation intelligent cyberinfrastructure for a computing continuum with heterogeneous resources that is usable in a plug-and-play manner by stakeholders to solve societal challenges?
Objectives: Intelligent CyberInfrastructure for Computing Continuum

Use Inspired Science Domains

Digital Agriculture

Smart Foodsheds

Animal Ecology

**ICICLE: Intelligent CyberInfrastructure with Computational Learning in the Environment**

Systems AI Foundational Research for CI

Emerging Computing Continuum

On Field Sensors

Edge & Near Edge

Clouds

HPC Systems & Data Centers

Cl for AI

AI for “Cl for AI”
Participation:
14 Organizations, 33 faculty, 41 staff, (58 PhD, 16 MS, 16 undergrad, 6 K-12) students & many Collaborators
Collaboration: ICICLE and the Technology Innovation Hub (TIH) at the Indian Institute of Technology Bombay (IIT-B), India

Digital Agriculture

This research collaboration will contribute novel design paradigms for context-adaptive CI and aims to develop next-generation CI for Digital Agriculture including AI and machine learning methods targeting 3 core areas.

Crop Health Modeling

- Sense crop health and level context to predict crop yield
- Detect stressors and diseases for geographically diverse crops
- Apply remedies with little human intervention via Internet of Things (IoT) and sensor systems

Aerial Crop Scouting

- CI for fully autonomous aerial systems
- Simplify deployment of UAV in real fields to capture common crop health conditions
- Provide accurate maps that yield valuable insights for crop management

Privacy-Preserving Data Exchange

Create secure, trustworthy, and privacy-preserving platforms that connect farmers and allow them to share information and resources safely.

Building upon the existing ICICLE infrastructure, CI and AI capabilities, researchers will leverage contextual conditions in India for Digital Agriculture that differ from the United States to (1) expose brittle CI components, (2) make AI4CI more robust and expansive in the long-term, (3) devise principles that yield context-aware CI.
Outline

• ICICLE Vision and Goals

• Research Challenges being Addressed
  • Overview of a few challenges

• Selected Accomplishments in the Digital Agriculture

• How to Get Engaged?
Research Plan: Overall Vision
Thrust: Foundational Systems AI

Components address CI complexity and heterogeneity for plug-and-play

- Knowledge Graphs
  - Multimodal KG to encode & reason rich data modalities (e.g., camera trap)
  - Auto construction
  - Interplay with LLM and knowledge-based QA

- Model Commons
  - MINT to support ICICLE use cases, KG, and models
  - Precise profiling
  - Flex composition
  - Versioning and provenance

- Adaptive AI
  - Context-aware
  - Efficient update
  - User-friendly adaptation process
  - Adaptation of foundation models, conversational AI

- Federated Learning
  - Heterogeneity
  - Context-aware
  - Privacy-preserving and robustness
  - Going beyond classification (GNN, foundation models)

- Conversational AI
  - KG- and model-commons-aware
  - LLM-powered
  - Grounding LLMs to the context
  - Hallucination reduction
  - Complex reasoning
Thrust: CI4AI

Provides necessary CI to deploy AI throughout computing continuum and make it plug-and-play!

High Perf. Training
- High-performance communication libraries
- Gradient sparsification
- Exploiting data-, model-, pipeline-, and hybrid-parallel paradigms

High Perf. Data Management
- Unified storage of data, model and hyperparameters
- Data location transparency with migration
- Leveraging new hardware

Edge Intelligence
- Performance characterization of edge
- Optimize ML/DL inference on edge devices
- Profiling edge devices to improve quality of service

AI-Adaptive Edge Wireless
- High-throughput, reliable communications
- Predictable Wireless Comm. via Rateless-Coding & Multi-Modal/Path
- AI-adaptive edge wireless prototypes.

Control and Coordination
- Functional/Performance Interface Design
- Intelligent Resource Management with Tapis
- Hardening and Optimizing for Production-ready Service

Thrust: CI4AI Provides necessary CI to deploy AI throughout computing continuum and make it plug-and-play!
Thrust: Privacy, Accountability and Data Integrity (PADI)

• PADI contributes to
  – ICICLE vision as transparent and trustworthy infrastructure for AI-enabled future
  – An ethically aligned infrastructure and workforce through an AI ethics framework

• PADI advances both technical and non-technical innovations and best practices that collectively contribute to a trusted environment
  – e.g., where stakeholders (farmers, industry partners, etc.) are comfortable contributing data and AI models for ICICLE AI research (and more broadly for AI research).

• PADI addresses both research questions and issues of practice (project norms and practice)
Co-Designing with use-inspired domains

Conv. AI

Data: On Field Sensors

Models: Edge & Near Edge

Data/Models: Cloud

Data/Models: HECs

ICICLE-enabled Computing Continuum
Outline

• ICICLE Vision and Goals
• Research Challenges being Addressed

• **Selected Accomplishment Highlights**
  • CI/Software Released
  • Digital Agriculture
    • HARVEST Service for Semi-Supervised Learning
    • Edge-to-Cloud Middleware for Aerial Crop Scouting
• How to Get Engaged?
CI/Software Components Released (so far)

2023.04 Release (04/30/23)

- **AI4CI**
  - HPC Application Runtime Predictor (HARP) v1.0
  - Intelligent Sparse Library (iSpLib) v1.0
- **Software and Reference Architecture**
  - Base ICICLE Tapis Software v1.3.0
  - Event Engine v0.2.0
  - Hello ICICLE Authentication Clients v0.0.1
  - Tapis Pods Service v1.3.0
  - CI Components Catalog v0.1.0
- **Animal Ecology**
  - Camera-Traps Edge Simulator v0.3.0
- **Digital Agriculture**
  - SoftwarePilot v1.2.5
- **Smart Foodsheds**
  - Persons-Projects-Organizations-Datasets (PPOD) Schema v0.9.1
  - Smart Foodsheds Visual Analytics (VA) Dashboard v0.1

2023.06 Release (06/30/23)

- **AI Foundations**
  - ICICLE Foodshed Parser v0.1
  - Species Classification using Multimodal Heterogeneous Context v0.1.0
  - Region2vec v1.0
- **Software and Reference Architecture**
  - Tapis Federated Authentication Service v1.3.4
  - ICICONSOLE v0.0.10
  - TapisCL-ICICLE v0.1.4
  - Tapis Pods Service v1.3.2
- **Animal Ecology**
  - Camera-Traps Edge Simulator v0.3.1
- **Digital Agriculture**
  - ICICLE Digital Agriculture Hub v1.0
  - Far-Edge Edge Simulator v1.0
  - In-Field Helper for Crop Scouts v1.0
- **Smart Foodsheds**
  - Persons-Projects-Organizations-Datasets_California (PPOD_CA) Knowledge Graph v23.06
  - Kroger Store Closure v0.1
  - Smart Foodsheds Visual Analytics (VA) Dashboard v0.2

[https://icicle.osu.edu/cyberinfrastructure/software](https://icicle.osu.edu/cyberinfrastructure/software)
CI/Software Components Released (so far)

2023.10 Release (10/06/23)

- AI4CI
  - HPC Application Runtime Predictor (HARP) v2.0
  - High Performance Computing Applications Dataset v1.0
- Software and Reference Architecture
  - iciflaskn v1.0
  - ICICONSOLE v0.8.0
  - TapisCL-ICICLE v1.0.11
- Animal Ecology
  - Camera-Traps Edge Simulator v0.3.2
- Smart Foodsheds
  - Smart Foodsheds Visual Analytics (VA) Dashboard v0.3

https://icicle.osu.edu/cyberinfrastructure/software
Outline

• ICICLE Vision and Goals
• Research Challenges being Addressed

• Selected Accomplishment Highlights
  • CI/Software Released
  • Digital Agriculture
    • HARVEST Service for Semi-Supervised Learning
    • Edge-to-Cloud Middleware for Aerial Crop Scouting

• How to Get Engaged?
ICICLE Use-Inspired Science: Digital Agriculture

Co-Leads

- Scott Shearer
  Food, Agriculture and Biological Eng.

- Christopher Stewart
  Computer Science & Eng

- Zichen Zhang
- Jenna Kline
  Ohio State University
- John C. Chumley
- Kevyn Angueira Irrizary

Digital Agriculture Hub and Use-Inspired Technologies

- P. Sadayappan
  University of Utah

- Jinghua Yan
  University of Utah

- Hari Subramoni
- Nawras Alnaasan

- Erman Ayday
  Case Western
- Beth Plale
  Indiana University
- Alfonso Morales
  University of Wisconsin

- Stakeholder Engagement

- Artificial Intelligence for Cyberinfrastructure
- Cyberinfrastructure for AI-Driven Digital Agriculture
- Privacy-aware, Explainable AI, & Democratization
The Application Domain Challenge (Digital Agriculture)

Computer Vision (CV) based classification scenarios are ubiquitous in use-inspired science domains such as Digital Agriculture.

- Residue Cover on Soil Surface
- Soil Aggregate Size
- Wheat Development
- Non-Uniform Emergence
- Nitrogen Deficiency
- European corn borer
- Corn leaf aphid
- Mexican bean beetle defoliation
The Data Labeling Challenge

- Data samples need to be fully labeled by an expert for training and evaluation.
- Datasets may be collected frequently and in large volumes (millions of unlabeled images).
- Labeling data by experts is a significant bottleneck.
- Supervised learning can be time-consuming, costly, and infeasible for certain applications.
The Computing Challenge

• Why do we need Parallel Training?

• Larger and Deeper models are being proposed
  – AlexNet -> ResNet -> NASNet – AmoebaNet -> ViT
  – DNNs require a lot of memory and a lot of computation
  – Larger models cannot fit a GPU’s memory

• Single GPU training cannot keep up with ever-larger models

• Community has moved to multi-GPU training

• Multi-GPU in one node is good but there is a limit to Scale-up (8-16 GPUs)

• Multi-node (Distributed or Parallel) Training is necessary!!
HARVEST (High-Performance ARtificial Vision Framework for Expert Labeling using Semi-Supervised Training)
Demo: Semi-Supervised Learning
Outline

• ICICLE Vision and Goals
• Research Challenges being Addressed

• **Selected Accomplishment Highlights**
  • CI/Software Released
  • **Digital Agriculture**
    • HARVEST Service for Semi-Supervised Learning
    • *Edge-to-Cloud Middleware for Aerial Crop Scouting*

• How to Get Engaged?
Digital Agriculture will transform crop management by:
   (1) sensing environmental conditions
   (2) characterizing crop health at fine granularities
   (3) autonomously delivering cost-effective treatments

Stakeholders include farmers and biologists—traditional agriculture professionals—and data scientists, machine learning experts, engineers, and HPC professionals.

ICICLE seeks to develop CI needed for all stakeholders to create, share, and process agricultural data effectively and efficiently.

In this context, AI will drive improvements in:
   (1) Autonomous, self-driving farms
   (2) Methods to wrangle the rapid growth of agricultural data
   (3) Data-driven and context-aware agricultural insights
   (4) Context-aware management and differential privacy
   (5) Managing open & democratized digital agriculture services
Digital Agriculture: Aerial Crop Scouting

• **Aerial Crop Scouting**: In this workload, we seek to create *heat maps* that describe crop health for a field
  - *Inform* self-driving tractors and sprayers to reduce the application of pesticide and fertilizer
  - *Predict* crop yields for harvest and market timing
  - *Identify* trends across farms, such as the introduction of resistant weeds

• **Technology**: Unmanned aerial vehicles (UAV) capture high resolution images
  - Flying low (15 ft above ground): 1 pixel -> mm

• **Transformative**: At mm-granularity, stakeholders can detect biological phenomena invisible to satellites
  - Soybean leaf defoliation caused by Japanese beetle

Courtesy of LaRue Farms Inc.
Demo: Cloud-to-Edge Middleware
Engagement with Other Organizations

• On-going discussions with several other AI Institutes
  • AIFARMS
  • AIIRA
  • AgAID
• Collaboration with TIH-Mumbai
  • More details will be provided tomorrow during US-India workshop
• Interactions with industry
  • CNH Industrial
  • TCS
Outline

• ICICLE Vision and Goals
• Research Challenges being Addressed
• Selected Accomplishment Highlights
• How to Get Engaged?
Multiple Levels of Collaboration and Engagement

- Using the Released Software/CI components
  - Available at [https://icicle.osu.edu/cyberinfrastructure/software](https://icicle.osu.edu/cyberinfrastructure/software)
  - Get engaged as a member in the Stakeholder Roundtable (more details below)

- Become a part of ICICLE (multiple options)
  - Student Associate
  - Visiting Research Fellow
  - Academic Collaborator
  - Industry Partner
  - Stakeholder Roundtable Member
  - More details at: [https://icicle.osu.edu/engagement/join-us](https://icicle.osu.edu/engagement/join-us)

- Join the ICICLE mailing lists ([https://icicle.osu.edu/engagement/mailing-lists](https://icicle.osu.edu/engagement/mailing-lists))
  - icicle-announce
  - icicle-discuss
Thank You!