

## ICICLE ANNOUNCES

### CI & SOFTWARE COMPONENTS RELEASE 2023-04

The ICICLE team aims to build the next generation Cyberinfrastructure (CI) to render Artificial Intelligence (AI) more accessible to everyone and to drive its democratization further in solving larger societal problems. It is with great pleasure that we announce the **first release of ICICLE CI components version 2023-04**. This release includes the following components:

#### Intelligent Cyberinfrastructure

- AI for CI-for-AI : HPC Application Runtime Predictor (HARP) v1.0
- AI for CI-for-AI : Intelligent Sparse Library (iSipLib) v1.0
- Software Architecture and Design : Base ICICLE Tapis Software v1.3.0
- Software Architecture and Design: Event Engine v0.2.0
- Software Architecture and Design: Hello ICICLE Authentication Clients v0.0.1
- Software Architecture and Design: Tapis Pods Service v1.3.0
- Software Architecture and Design: CI Components Catalog v0.1.0

#### Use Inspired Science

- 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

The ICICLE team is committed to delivering the best software and CI components. We welcome your feedback and suggestions for future releases. A list of all ICICLE components can be found on our CI & Software [webpage](https://icicle.osu.edu/cyberinfrastructure/software) (<https://icicle.osu.edu/cyberinfrastructure/software>)

#### Acknowledgements

This release is brought to you by the National Science Foundation (NSF) funded AI institute for Intelligent Cyberinfrastructure with Computational Learning in the Environment (ICICLE) (OAC 2112606)



## MODELING A FRAMEWORK TO ESTIMATE HPC RESOURCE REQUIREMENTS FOR SCIENTIFIC WORKFLOWS USING HARP (HPC APPLICATION RESOURCE PREDICTOR)

Researchers use high-performance computing (HPC) cyberinfrastructures (CI) like Ohio Supercomputer (OSC) or Texas Advanced Computing Center (TACC) to execute computationally intensive diverse scientific workflows. Some workflows are heavy on IO, like genome sequencing (cleaning and assembly), while others, like training DNNs, could be compute (and memory) intensive. Each workflow has a unique resource requirement, and it is essential to profile and understand these needs to allocate shared resources for optimal utilization of the cyberinfrastructure. These resources are expensive, and several jobs compete to get these allocations, sometimes with reasonable wait times (while requesting enormous resources for a long time). Estimating the expected resources for optimally utilizing the compute and memory is challenging especially considering the need for sufficient history to enable these predictions tailored for unique workflows and execution environments. We explored and established a framework (as shown in Figure 1) that pipelines the solutions to address these challenges. The framework is configured to generate a history of executions and train suitable regression models to estimate the approximate execution time for a targeted application.

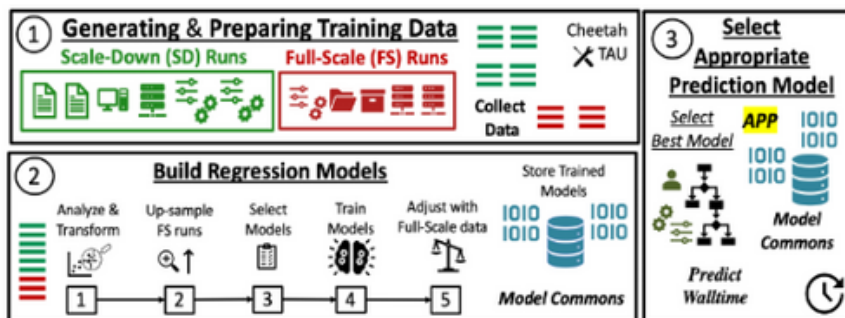


Fig 1: The Proposed Framework: training data generation, building regression models & selecting the best model based on custom criteria

Components and Characteristics of the framework (from figure 1):

- **Generating and Preparing Training Data:** This module automatically and systematically generates comprehensive, diverse "scaled-down(SD)" and limited, selective "full-scale(FS)" runs with minimal human intervention. We use Cheetah (<https://github.com/CODARcode/cheetah>) to execute the target application with the pre-defined data generation configurations (SD and FS) to generate the history-of-runs training data.
- **Building Regression Models:** This module standardizes and prepares the data, trains the selected off-the-shelf regression models with the appropriate hyper-parameters, and stores them for inference. In this phase, the data generated in the first phase is processed to train regression models. Redundant features are eliminated, outliers are removed, and features are transformed to reduce the dimensionality before training the regression models.
- **Selecting Appropriate Prediction Model:** this module selects the most appropriate regression model from a pool of trained models from phase 2 with respect to a given policy and target application.

Download Instructions are available here <https://github.com/ICICLE-ai/harp>.

For additional information please contact S. Vallabhajosyula at

vallabhajosyula.2@buckeyemail.osu.edu



## **STRENGTHENING US-INDIA COLLABORATION ON DIGITAL AGRICULTURE (TIH-NSF SUPPLEMENT GRANT)**



Dr. Christopher Stewart, Associate Professor, Computer Science and Engineering at The Ohio State University traveled to India to meet with ICICLE collaborators in Delhi and Mumbai. These collaborations were formed through our TIH-NSF supplement grant titled "Computational Learning through Context Adaptation for Effective and Efficient Agriculture", one of the 35 joint research and development projects supported by National Science Foundation (NSF) & Technology Innovation Hubs (TIHs) supported by the Department of Science and Technology (DST), Government of India under the National Mission on Interdisciplinary Cyber-Physical Systems.

He gained a much richer perspective of agriculture challenges in the Indian context during this visit. For example, the demand for research on fully autonomous vehicles is lower in India compared to the United State because farm labor is more readily available. According to Dr. Stewart, "Our collaborators also have rich, custom data sensing units in the ground at various sites across India, a mechanism we had not yet considered in our collaboration".

In Mumbai, he was hosted by Professors Rajbabu Velmurugan, Maryam Shojaei, Arpita Sinha and their student Rugved Katole. Dr. Stewart discussed immediate applications for aerial scouting units that require multiple concurrent AI models for use cases such as plant detection, soil health, and crop variety. New problems were discovered because of a week-long immersion between the groups. They compliment existing problems such as transfer learning, privacy, and aerial-image data collection. All of these problems are important to crystallize.

In Delhi, Dr. Stewart met with Ajay Gupta of Aiesky Advisory & Consultancy Services, a startup using satellite imagery to digitize and visualize farmland ownership and operations throughout India. They discussed advancing Aiesky's existing cyber infrastructure to capture field conditions by integrating reinforcement learning packages (see MARbLE, SEC 2022).



## **EXCITING NEWS: SUCCESSFUL LAUNCH OF THE ICICLE NEXTGENS STUDENT COMMUNITY**

On April 27th, the NextGens Student Community held a virtual kickoff meeting to introduce itself as a value-add that is an automatic benefit to all students involved in ICICLE regardless of age. As was reviewed in the session, the majority of students are at the graduate level, however, there are participating students from high school through PhD candidates. The session was a casual event where student attendees were given time to meet each other and voice feedback and opinions of what they'd like to see in the near future regarding the community. The general consensus was positive that if there is a way to connect and share what students are working on across projects that it may be beneficial to all as everyone is a part of the bigger picture of ICICLE. The community is run by students and is backed by a panel of ICICLE professionals who would like to support the students with any questions, etc.

Currently Victoria De Bella, 2nd year MBA in Marketing Analytics & Insights at the University of Wisconsin-Madison helped launch the NextGens Student Community, but as she is graduating soon we are looking for other ICICLE students who would be interested in helping continue the community.

Please contact Victoria directly at [debella@wisc.edu](mailto:debella@wisc.edu) with any questions. Good luck to all students with their final exams and to any graduates! Be on the lookout for upcoming events from the NextGens Student Community or feel free to ask a question in the designated ICICLE Teams channel!

## **ICICLE'S TIPS ON ALLYSHIP**

The Broader Impact Backbone Network team sends out tips in video format on how we, here at ICICLE, can become better allies

**TO PROMOTE AN AWARE, INCLUSIVE, AND MORE DIVERSE COMMUNITY**

**Browse all Tips for Allyship [here](#)**



## **RECENT PRESENTATIONS & PUBLICATIONS**

### **PRESENTATIONS**

- Song Gao, "Convergence of CyberGIS and Geospatial AI", In the AAG 2023 Symposium on Harnessing the Geospatial Data Revolution for Sustainability Solutions Panel Session, March 2023
- Yu Su, "Grounding Language Models to Real-World Environments", University of Tokyo, March 2023.

### **PUBLICATIONS**

- Sahil Samar, Michael Ray, James Karpinski, Mia Chen, Archita Sarin, Christian Garcia, Matthew Lange, Joe Stubbs, Mary Thomas, "Development of Authenticated Clients and Applications for ICICLE CI Services - Final Report for the REHS Program", April 2023, [PDF]
- Yamei Tu, Xiaoqi Wang, Rui Qiu, Han-Wei Shen, Michelle Miller, Jinmeng Rao, Song Gao, Patrick R Huber, Allan D Hollander, Matthew Lange, Christian R Garcia, Joe Stubbs," An Interactive Knowledge and Learning Environment in Smart Foodsheds", IEEE Computer Graphics and Applications, April 2023. [PDF]

Click [here](#) to view ICICLE Presentations and Publications



**ICICLE**  
DEMOCRATIZING AI

## SUBSCRIBE TO ICICLE MAILING LISTS

The following mailing lists are available for ICICLE software and cyberinfrastructure releases, future updates and miscellaneous questions regarding installation/build problems, performance issues.

- **icicle-announce**: This is an announcement list only. If you would like to get information about future updates, software and cyberinfrastructure releases, publications, etc. related to the ICICLE project, you may subscribe to this mailing list. This list is open to public. You are welcome to subscribe to this mailing list yourself.
- **icicle-discuss**: This is a discussion list. This mailing list is meant for users to discuss all installation/build problems, performance issues , features and other miscellaneous questions related to the different software and cyberinfrastructure components of the ICICLE project. In order to post your questions and suggestions to this mailing list, you need to be a registered user of ICICLE with an organizational e-mail address and be a member of this list by subscribing to it with the same e-mail address. If you are not a registered user, please follow the procedure indicated under Download tab in the top menu to get registered.

We welcome your interest to partner with ICICLE!  
Please complete this **form** and we'll reach out to you.



Follow us @icicleai



Newsletter [Archives](#)

