



ABE Spring Faculty Seminar Series

Somali Chaterji

Assistant Professor
Dept. of Agricultural & Biological Engineering

Tuesday, March 3, 2020
1:30 - 2:30 p.m. - BROWN 1154

"Artificial Intelligence, Scalable Data Lakes, and On-Drive Computation: Applications in Digital Agriculture and Computational Genomics"

Abstract:

We live in a data-driven world! Work in my laboratory *ICAN* is motivated by the vision to leverage data to build transformative algorithms for propelling advances in data-driven agriculture and health. In the **digital agriculture thrust** of my work, we believe in harnessing the power of small, seemingly computationally frail, embedded devices to run compute-heavy applications such as neural network-driven computer vision applications. How can we make these embedded devices handle the heavy load and yet execute in real-time and with a low energy budget? To answer questions like these, *ICAN* is developing systems for on-device computation for "onloading" computation on the device, which can preserve privacy and security, and minimize data transfer overheads when offloading computation to the cloud or to the edge. Another aspect is how to architect robust data lakes for workload-aware and availability-aware data storage and access. These databases can "live" on premise or, as is often the case, be hosted on the cloud. How then can the databases be made to deliver peak performance, adapting to the variable workloads in a typical IoT framework, such as those envisioned by digital agriculture applications. The second thrust of my talk will be on **AI building blocks for genome engineering**. In this part, I will present our use of Natural Language Processing (NLP) techniques for efficient genomic error correction, then present a data-parallel framework for running intensive computational genomics applications, and finally, a Domain Specific Language (DSL) that can be used to create modular new applications or to evolve existing applications. Conclude the talk with discussion of the interplay of the two thrusts in my innovatory and a plug for the new series of stackable and applied data science courses I am developing for applied digital agriculture and biological engineering.

Bio:

She works in the areas of data science and engineering for digital agriculture and genomics and in building cyberinfrastructures for these domains. She is a part of the WHIN (Wabash Heartland Innovation Network) inspiration team working to bring IoT to advance agriculture. Data is power! She believes in leveraging the power of big data to make digital agriculture efficient and for predictive and personalized healthcare, one in which health and vitality is the focus. Dr. Chaterji got her PhD in Biomedical Engineering from Purdue University, winning the Chorafas International Award (2010), College of Engineering Best Dissertation Award (2010), and the Future Faculty Fellowship Award (2009). She did her Post-doctoral Fellowship at the University of Texas at Austin in the Department of Biomedical Engineering, where her work was supported by an American Heart Association (AHA) award (2014). She has won the best paper award at the ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB) conference (2015). Dr. Chaterji is also a lab-to-bedside commercialization enthusiast and is a scientific advisor to the IC2 Institute at the University of Texas at Austin since 2014 and recently filed two patent applications based on her work on scalable data lakes. Dr. Chaterji won Purdue's Seed-for-Success Award for winning a research grant of more than \$1M (NIH R01, 2016) to improve the cyberinfrastructure for computation for the microbiome, specifically for the world's largest microbiome repository and computation engine--MG-RAST housed in the Argonne National Laboratory.