

**University of California, Irvine  
Statistics Seminar**

***Fast and Powerful Minipatch Ensemble Learning  
for Discovery and Inference***

**Genevera Allen  
Associate Professor  
Rice University**

**4 p.m., Thursday, April 13, 2023  
6011 Donald Bren Hall**

Enormous quantities of data are collected in many industries and disciplines; this data holds the key to solving critical societal and scientific problems. Yet, fitting models to make discoveries from this huge data often poses both computational and statistical challenges. In this talk, we propose a new ensemble learning strategy primed for fast, distributed, and memory-efficient computation that also has many statistical advantages. Inspired by random forests, stability selection, and stochastic optimization, we propose to build ensembles based on tiny subsamples of both observations and features that we term minipatches. While minipatch learning can easily be applied to prediction tasks similarly to random forests, this talk focuses on using minipatch ensemble approaches in unconventional ways: making data-driven discoveries and for statistical inference. Specifically, we will discuss using this ensemble strategy for feature selection, clustering, and graph learning as well as for distribution-free and model-agnostic inference for both predictions and important features. Through huge real data examples from neuroscience, genomics and biomedicine, we illustrate the computational and statistical advantages of our minipatch ensemble learning approaches.

Genevera Allen is an Associate Professor of Electrical and Computer Engineering, Statistics, and Computer Science at Rice University and an investigator at the Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital and Baylor College of Medicine. She is also the Founding Director of the Rice Center for Transforming Data to Knowledge, informally called the Rice D2K Lab.

Dr. Allen's research develops new statistical machine learning tools to help people make reproducible data-driven discoveries. She is known for her work in the areas of interpretable machine learning, data integration, modern multivariate analysis, and graphical models with applications in neuroscience and bioinformatics. Dr. Allen is also a leader in data science education. In 2018, she founded the Rice D2K Lab, a campus hub for experiential learning and data science education. Through her leadership of the D2K Lab, Dr. Allen developed new interdisciplinary data science degree programs, established a novel capstone program in data science and machine learning, and led Rice's engagement with corporate and community partners in data science.