

## Automating Machine Learning Model Building with Big Clinical Data

**Gang Luo, Ph.D.**  
Associate Professor  
Biomedical Informatics and Medical Education  
School of Medicine  
University of Washington

**Abstract:** Predictive modeling is fundamental for extracting value from large clinical data sets, or “big clinical data,” advancing clinical research, and improving healthcare. Predictive modeling can facilitate appropriate and timely care by forecasting an individual’s health risk, clinical course, or outcome. Machine learning is a major approach to predictive modeling using algorithms improving automatically through experience, but two factors make its use in healthcare challenging. First, before training a model, the user of a machine learning software tool must manually select a machine learning algorithm and set one or more model parameters termed hyper-parameters. The algorithm and hyper-parameter values used typically impact the resulting model’s accuracy by over 40%, but their selection requires special computing expertise as well as many labor-intensive manual iterations. Second, most machine learning models are complex and give no explanation of prediction results. Nevertheless, explanation is essential for a learning healthcare system.

To automate machine learning model building with big clinical data, we are currently developing a software system that can perform the following tasks in a pipeline automatically:

- (a) select effective machine learning algorithms and hyper-parameter values to build predictive models;
- (b) explain prediction results to healthcare researchers;
- (c) suggest tailored interventions; and
- (d) estimate outcomes for various configurations, which is needed for determining a proper strategy to deploy a predictive model in a healthcare system.

This talk will present the design, initial implementation, and some preliminary results of the software system.

**Bio:** *Gang Luo* obtained his Ph.D. degree in Computer Science minor in Mathematics at the University of Wisconsin-Madison in 2004. Between 2004 and 2012, he was a Research Staff Member at the IBM T.J. Watson research center. Between 2012 and 2016, he was a faculty member in the Department of Biomedical Informatics at the University of Utah. Gang is currently a faculty member in the Department of Biomedical Informatics and Medical Education of the School of Medicine at the University of Washington. His research interests include health/clinical informatics (software system design/development and data analytics), big data, information retrieval, database systems, and machine learning with a focus on health applications. He invented the first method for automatically providing rule-based explanation for any machine learning model's prediction/classification results without degrading accuracy, the questionnaire-guided intelligent medical search engine iMed, intelligent personal health record, and SQL and compiler progress indicators.

**Tuesday, March 28, 2017**  
**1:30 – 2:20 p.m.**  
**MEB 235**