



## Machine Learning and Data Mining

Rapidly-growing “Big Data” resources have posted many interesting new challenges for machine learning, with the data being large, loosely labeled, highly diverse, complex and often relationally structured. We have proposed a diverse range of machine learning approaches, including [feature learning](#), [deep learning](#), [sparse modeling](#), [latent factor analysis](#), semi-supervised learning, multi-task learning, and others, to handle different types of data complexities that are urgent to be addressed in this domain. We strive toward building and sharing benchmarked datasets and open-source releases of research prototypes.

## Applications in Biomedicine and More

My group takes an inter-disciplinary approach, combining ideas from statistical machine learning, computational biology, computer architecture, secure computation and language technologies. Over years we have had valuable opportunities to work on many practically important applications, covering multiple research fields, such as proteomics, cancer genomics, medical informatics, bio-text mining, structural biology, immunology, more traditional text mining topics, like information extraction, ad hoc information retrieval and text/image/video labeling and recently on cloud system analysis and sensor data mining.

## RECENT RESEARCH DEVELOPMENTS

- Beilun Wang, Ji Gao, Yanjun Qi, (2017) "A Fast and Scalable Joint Estimator for Learning Multiple Related Sparse Gaussian Graphical Models", Proceedings of The 20th International Conference on Artificial Intelligence and Statistics (AISTATS)
- Ritambhara Singh, Jack Lanchantin, Yanjun Qi, (2016) "DeepChrome: Deep-learning for predicting gene expression from histone modifications", 15th European Conference on Computational Biology, (ECCB-16) (Bioinformatics (2016) 32 (17): i639-i648.)
- Zeming Lin, Jack Lanchantin, Yanjun Qi, (2016) "MUST-CNN: A Multilayer Shift-and-Stitch Deep Convolutional Architecture for Sequence-based Protein Structure Prediction", (AAAI)
- Jack Lanchantin, Ritambhara Singh, Zeming Lin, Yanjun Qi, (2016) "Deep Motif: Visualizing Genomic Sequence Classifications", the International Conference on Learning Representations (ICLR)

## RECENT AWARDS

- NSF CAREER Award (2015-2020)
- Best Paper Award @ International Conference on Body Area Networks (BodyNets) – 2014
- Text Retrieval Conference (TREC) Medical Record Retrieval Competition, Second Place - 2012

## SEAS Research Information

Pamela M. Norris,  
Executive Associate Dean for Research  
University of Virginia  
Box 400232  
Charlottesville, VA 22903  
[EngineeringResearch@virginia.edu](mailto:EngineeringResearch@virginia.edu)  
434.243.7683



SCHOOL of ENGINEERING  
& APPLIED SCIENCE