Yichao Li

Stocker Center 311, School of Electrical Engineering and Computer Science, Ohio University

Tel: 740-591-1289 Email: YL079811@ohio.edu   Website: https://YichaoOU.github.io

Education

Ph.D.	2018	Ohio University, Electrical Engineering and Computer Science, GPA: 3.9
		Dissertation title: Algorithmic Methods for Multiomics Biomarker Discovery
		Expected graduation: December 2018
MS	2017	Ohio University, Mathematics
MS	2015	Ohio University, Computer Science
BS	2008	Capital Normal University (Beijing, China), Biological Science
BS	2008	Capital Normal University, Computer Science

Machine Learning Competitions

2018	Kaggle silver medal, Top 5 %, 18/394, The 2nd YouTube-8M Video Understanding Challenge
2018	3rd place, Autonomous Vehicle Drag Race 2, IARRC 2018
2017	2nd place, AI in cancer prognosis, Multiple Myeloma DREAM Challenge SC 1
2017	Top 18%, 660/3779, Zillow Prize: Zillow's Home Value Prediction (Zestimate), Kaggle
2016	7th place, 7/34, Computational Genomics, the CTCF group, the ENCODE-DREAM Challenge

Machine Learning Project

2018	Founder, **Triplan Alpha**, ongoing work, demos available upon request
	An AI model to efficiently and effectively schedule the best possible trips/vacations using existing data from hotel, restaurant, and POI reviews, i.e. APIs of Google Places, Yelp, TripAdvisor.

Patent

2017	R Hoffmann, A Frolov, S Spiller, **Y Li**, LR Welch. Method and means for the non-invasive diagnosis of type ii diabetes mellitus. US20170045533A1.

Research Experience

2012 - Present	Research Assistant, Ohio University Bioinformatics Laboratory
	Projects: Characterization of 3D genome organization, Novel machine learning models for protein-DNA interactions and human disease studies (e.g. cancer, diabetes)
2011 - 2012	Data Scientist, Beijing Computing Center

Research Interests

Machine Learning, Statistics, Healthcare Informatics, Deep Learning, Predictive Analytics, Data Visualization, Data Mining, Computer Vision, Text Mining

Big Data Club

I started the Big Data Club, an OU student organization, with the goal to provide learning activities for data analytics. I am excited and proud to serve the Athens community and help OU Bobcats develop essential skills for data science. Through international competitions, my efforts will improve the national standing for OU.

Teaching Experience

2018	President, Big Data Club, Ohio University

Designed and taught 15 data science tutorials.
Helped students to develop core data analytic skills using Python.

2013 - 2017 Teaching Assistant, Ohio University

1. CS4900/CS5900/CS6900, 8 semesters, Bioinformatics Journal Club
   - Worked with a diverse population of students, including many international students.
   - Helped to select journal articles.
   - Encouraged students to participate in class discussions.

2. CS4170/CS5710, 4 semesters, Data Mining With Applications in the Life Sciences
   - Designed and hosted in-class bioinformatics competitions.
   - Helped to solve students' in-class questions, e.g. Perl, R, MySQL
   - Invited to teach 2 lectures of data analytic techniques and machine learning algorithms
   - Helped to design and grade final machine learning project

3. CS4160/CS5160, 4 semesters, Bioinformatics Tools
   - Invited to teach RESTful API, motif discovery, and machine learning algorithms
   - Graded various types of assignments, e.g. codes, projects, quizzes, research papers.
   - Encouraged students to participate in class discussions.

4. CS2400, 2 semesters, Lab sessions, Intro to Computer Science I
   - Taught C programming and Linux.

2016 Free lectures: Building personal websites on GitHub using Jekyll.

Attracted more than 25 faculty and students.
Designed and taught 4 lectures. https://yichaoou.github.io/tutorials/web_design/

2012 - 2013 Grading Assistant, Ohio University

2 semesters, Applied Probability and Statistics for Electrical Engineers

Teaching Interests

Data Mining, Data Visualization, Predictive Analytics, Machine Learning, Deep Learning, Theory of Statistics, Applied Statistics, Algorithms, Advanced Algorithms, Computation Theory, Software Engineering, Programming Languages, Data Structure, Databases, Computer Graphics, Parallel Processing

Awards

2018 1st place, Computational Biology/ISCB Wikipedia Writing Competition
https://en.wikipedia.org/wiki/Chromosome_conformation_capture

2017 2nd place, Ohio University Student Expo poster competition

Professional Services

2018 Professional Autonomous Vehicle Engineers, Dr. Jim Zhu lab, Ohio University
Developed a computer vision model on NVDIA TX1 and TK1 using TensorFlow and OpenCV.

2017 Student speaker, Ohio University EECS Advisory Board Retreat
Panel discussion of personal experiences

2016 Reviewer, PLOS Computational Biology

Publications

Google Scholar Profile: https://scholar.google.com/citations?user=Rm_GiJYAAAAJ&hl=en

Journal Papers

1. S Spiller, **Y Li**, M Blüher, L Welch, R Hoffmann. Diagnostic Accuracy of Protein Glycation Sites in Long-Term Controlled Patients with Type 2 Diabetes Mellitus and Their Prognostic Potential for Early Diagnosis. *Pharmaceuticals*. 2018
2. K Y Lee, R Sharma, G Gase, S Ussar, **Y Li**, L Welch, D E Berryman, A Kispert, M Blüher, and C R Kahn. Tbx15 Defines a Glycolytic Subpopulation and White Adipocyte Heterogeneity. *Diabetes*. 2017
3. S Spiller, **Y Li**, M Blüher, L Welch, R Hoffmann. Glycated lysine-141 in haptoglobin improves the diagnostic accuracy for type 2 diabetes mellitus in combination with glycated hemoglobin HbA1c and fasting plasma glucose. *Clinical proteomics*. 2017.
4. **HPN-DREAM Consortium**. Inferring causal molecular networks: empirical assessment through a community-based effort. *Nature methods*. 2016.

Work-in-progress Papers

5. **Y Li**, Y Liu, D Juedes, F Drews, R Bunescu, L Welch. Set Cover Based Methods for Motif Selection in ChIP-seq Data. Submitted to *Bioinformatics*. 2018.
6. **Y Li**, X Liang, F Drews, L Welch, L Elnitski. Transcriptome sequencing and promoter analysis in papillary serous and low malignant potential ovarian tumors. To be submitted to Gynecological Oncology. 2018.
7. **Y Li**, A Showalter, D Juedes, L Welch. A Novel Discriminative Set Multi-Cover Model for Discovering Co-occurring DNA Motifs. To be submitted to *IEEE/ACM TCBB*. 2018.

Conference Paper

8. **Y Li**, XA Shen, RL Ewing, J Li. Terahertz spectroscopic material identification using approximate entropy and deep neural network. *Proceedings of 2017 IEEE National Aerospace & Electronics Conference.*

Project Experience

Pipeline Automation

* Developed and parallelized numerous data analytic pipelines using Python, R, and Bash, and applied it to multiple human diseases (e.g. ovarian cancer, breast cancer) and cell types. Deployed all pipelines to the Owens cluster at the Ohio Supercomputer Center.

Machine Learning

* Developed a diagnostic model for diabetes that integrated mass spectrometry data, clinical data, and demographic data using decision tree, SVM-RFE, random forest-RFE, k-means, and EM.
* Developed a linear stacking model (including XGBoost) using whole-genome exon sequencing data and RNA-seq gene expression data.
* Developed a random forest model to predict *in vivo* binding of 32 transcription factors in 13 cell types.

Information Retrieval

* Developed DNA motif ranking algorithms using tf-idf.
* Introduced a novel set cover problem (Boolean Retrieval) and proved its NP-completeness. Developed an approximation algorithm and proved its theoretical properties.

Deep Learning

* Developed a cancer classification model for 6 different types based on stacked sparse autoencoder using 791 microarray datasets.

TensorFlow

* Characterized 3D genome organization and developed whole-genome analysis algorithm using GPU.

Full-stack Web Development

- DNA barcoding analysis: http://bcat.cs.ohio.edu.

Big Data Analytics

- Hands-on experience with Spark, Hadoop, Google Cloud, MangoDB.

Optimization Algorithms

- Systematically evaluated several set cover algorithms using nested cross-validation, including genetic algorithms, tabu search, and relaxed integer linear programming.

Statistics

- Applied hypothesis testing in numerous projects, including ANOVA, t-test, Fisher exact test, Chi-square test, and power analysis.
- Identified markers of glycolytic metabolism using multiple linear regression analysis from qPCR data.
- Identified cell distributions in adipose tissues using kernel density estimation.

Network Analysis

- Built a global gene co-expression network in Barley and identified modules using network clustering.
- Built a chromatin interaction network and calculated network statistics, e.g. betweenness centrality.

Contributed Talks

| 2017 | Reverse Engineering of The Human Genome: Predicting Protein-DNA Interactions Across Multiple Cell Types, *Appalachian Region Cell Conference* |

Poster Presentations

2018	Barley Global Gene Co-expression Network Construction using RNA-Seq data, Ohio University Student Expo
2018	Parallelization of an optimized DNA motif analysis pipeline, with applications in multiple species, *Statewide Users Group meeting at the Ohio Supercomputer Center*
2017	Reverse Engineering of The Human Genome: Predicting Protein-DNA Interactions Across Multiple Cell Types, *Statewide Users Group meeting at the Ohio Supercomputer Center*
2017	Reverse Engineering of The Human Genome: Predicting Protein-DNA Interactions Across Multiple Cell Types, *ISCB Great Lakes Bioinformatics Conference*
2017	Reverse Engineering of The Human Genome: Predicting Protein-DNA Interactions Across Multiple Cell Types, Ohio University Student Expo
2017	Kernel Classification and Visualization: Unfold the Art of Kernel Trick, Mathematics Department's annual Pi-day, Ohio University
2016	Discovering Gene Regulatory Elements Using Coverage-based Heuristics, *Cold Spring Harbor conference on Systems Biology*
2016	Motif Discovery in co-regulated DNA sequences, *ISCB Great Lakes Bioinformatics Conference*
2016	Motif Discovery in co-regulated DNA sequences, Ohio University Student Expo
2016	Data Visualization: Unfold the Art of Data Mining Algorithms, Mathematics Department's annual Pi-day, Ohio University
2015	Epigenetic information improves genome-wide motif discovery, *ISCB Great Lakes Bioinformatics Conference*
2015	Epigenetic information improves genome-wide motif discovery, Ohio University Student Expo

2015	Deep Learning: Stacked Sparse Auto-encoder, Mathematics Department's annual Pi-day, Ohio University
2014	Homology modeling of extensin peroxidase in Solanum Lycopersicum, *ISCB Great Lakes Bioinformatics Conference*
2014	Homology modeling of extensin peroxidase in Solanum Lycopersicum, Ohio University Student Expo