

# Jinghe Zhang

---

**Contact** Olsson Hall 227C, 151 Engineer's Way, Charlottesville, VA 22903

**Information** Homepages:  
<https://www.linkedin.com/in/jinghezhang>  
<https://github.com/JingheZ>  
<http://people.virginia.edu/~jz4kg/>  
Phone: 607-651-8906 | Email: jz4kg@virginia.edu

**Research** **Machine Learning**, e.g., representation learning, metric learning;  
**Interests** **Machine Learning Applications**, e.g., data mining, predictive modeling, recommender systems, health and medical informatics

**Education** **University of Virginia** *08/2013 - 05/2017 (expected)*  
Ph.D., Systems & Information Engineering *Charlottesville, VA*  
GPA: 3.91/4.0

**Binghamton University, State University of New York** *08/2011 - 05/2013*  
Master of Science, Industrial & Systems Engineering *Binghamton, NY*  
GPA: 3.76/4.0

**Hebei University of Technology** *09/2007 - 06/2011*  
Bachelor of Science, Industrial Engineering *Tianjin, China*  
GPA: 3.65/4.0

**Technical** **Programming:** Python, Java, R, Matlab, SQL, HTML, CSS, Hadoop, Spark  
**Skills** **Operating Systems:** Windows, Unix/Linux  
**Tools:** Latex, Git, SVN, Vim, Emacs, Arena, Simio, ExpertFit, Minitab, NetLogo  
**Languages:** English and Chinese

**Experience** **Graduate Research Assistant** *08/2013 - Present*  
Department of Systems & Information Engineering, University of Virginia

**High-risk Patients Retrieval using Electronic Health Data** *02/2015 - Present*

- Perform representation learning on patients' medical histories to construct feature vectors with reduced information loss than using common methods
- Conduct metric learning on the representation of patients' medical data and discovered informative features for clinical use, such as transitions from unrelated diagnoses of anxiety/depression to potentially related ones
- Train well-established classifiers, such as support vector machine, linear discriminant analysis, random forest, k nearest neighbors, etc., to discriminate patients with and without anxiety/depression
- Evaluate the proposed framework by the early detection of anxiety/depression on new patients and the preliminary results achieved a 1%–4.5% higher accuracy compared to baseline methods
- Communicate with professors and MDs to define research problems, to explore potential methodologies, and to gather feedbacks regularly from them

**Predictive Modeling on Hyperlactatemia Sepsis** *05/2015 - Present*

- Analyze the characteristics of different ICU patient populations extracted from the MIMIC-II database, including mortality rates, the usage of antibiotics and intravenous fluids, and timing of severe sepsis, etc.
- Examine different definitions of severe sepsis and its correlation with ICU, hospital, and 30-day mortalities
- Build predictive models on a patient's likelihood of developing severe sepsis using statistical approaches (e.g., Cox proportional hazards model and logistic regression) and machine learning algorithms (e.g., support vector machine and random forest)
- Predict the real-time probability of lactate clearance considering a patient's demographic and physiological characteristics and clinical interventions
- Teamworked with professors, MD, data scientists, and other colleagues from medical and engineering backgrounds and lead weekly meetings to deliver presentations and to conduct discussions

## Text Mining on Restaurant Reviews from Yelp

03/2015 - 04/2015

- Preprocessed 1 million restaurant reviews crawled from Yelp with tokenization, stemming, stop word removal, and normalization, to construct N-gram vector space representation for text documents and computed similarities among different documents
- Implemented statistical language models with maximum likelihood estimation and smoothing; generated text documents from language models and evaluated the constructed N-gram language models
- Developed a text categorization system, including feature selection, Nave Bayes and KNN classifier, to distinguish positive and negative restaurant reviews
- Evaluated the text categorization system with 10-fold cross-validation and performed parameter tuning to explore the best configuration of KNN with brute force and random vector projection, respectively.

## Measuring Care Variation among Congestive Heart Failure Patients

09/2013 - 08/2014

- Applied vector space models to measure the similarities between congestive heart failure patients based on their medication and procedure orders
- Utilized clustering (K-means and co-clustering using bipartite graph) and topic modeling methods to detect patient subgroups within the congestive heart failure population according to medication orders, diagnostic and demographic information
- Teamworked with software engineers to implement the above models using Spark on Hadoop clusters with significantly decreased computation time

## Web Crawler and Document Analysis & Retrieval

09/2014 - 11/2014

## Spam Filtering Using Machine Learning Approaches

09/2013 - 11/2013

## Graduate Teaching Assistant

01/2014 - 05/2014

SYS6016 Machine Learning, University of Virginia

## Graduate Research Associate

08/2011 - 05/2013

Watson Institute for Systems Excellence, Binghamton University  
Quality Management Department, United Health Services (UHS)

## Predictive Modeling on 30-day Hospital Readmission

04/2012 - 05/2013

- Collected clinical and demographic data and identified significant risk factors in readmission using statistical analysis
- Accomplished predictive models using Naïve Bayes, particle swarm intelligence-based support vector machine (SVM), random forest, and ensemble neural networks, and improved average prediction accuracy by 22% to assist clinicians in identifying high-risk patients

## Selected Publications

1. **Jinghe Zhang**, Haoyi Xiong, Yu Huang, Hao Wu, Kevin Leach, and Laura E. Barnes. *M-SEQ: Early Detection of Anxiety and Depression via Temporal Orders of Diagnoses in Electronic Health Data*, 2015. *IEEE International Conference on Big Data (IEEE BigData 2015)*. Oct 29 - Nov 1. Santa Clara, CA.
2. Bichen Zheng, **Jinghe Zhang**, Sang Won Yoon, Sarah S. Lam, Mohammad Khasawneh, and Srikanth Poranki. Predictive modeling of hospital readmissions using metaheuristics and data mining, 2015. *Expert Systems With Applications* vol.42, no.20, pp 7110-7120.
3. **Jinghe Zhang**, Sarah S. Lam, and Srikanth Poranki. A Classification Model for Hospital Readmission Using Combined Neural Networks, 2013. *Proceedings of Industrial and Systems Engineering Research Conference (ISERC)*. May 18-22. San Juan, PR.
4. **Jinghe Zhang**, Sang Won Yoon, Mohammad Khasawneh, Srikanth Poranki, and Krishnaswami Srihari. A Readmission Prediction Model Using Swarm Intelligence-based Support Vector Machine, 2013. *Proceedings of Industrial and Systems Engineering Research Conference (ISERC)*. May 18-22. San Juan, PR.
5. Chanchal Saha, **Jinghe Zhang**, Sang Won Yoon, Mohammad Khasawneh, and Krishnaswami Srihari. Selection and Matching of Kidney Donor and Recipient Using Fuzzy Techniques and Analytic Hierarchy Process, 2012. *Proceedings of Industrial and Systems Engineering Research Conference (ISERC)*. May 18-22. Orlando, FL.

## Honors & Awards

Graduate Research Scholarship, University of Virginia, 2014-2016  
Commonwealth Fellowship, University of Virginia, 2013-2014  
Graduate Research Scholarship, Binghamton University, 2011-2013  
National Aspiration Scholarship, China's Ministry of Education, 2009