

## Jinghe Zhang

---

CONTACT INFORMATION	Olsson Hall 227C, 151 Engineer's Way, Charlottesville, VA 22903 Homepages: <a href="https://www.linkedin.com/in/jinghezhang">https://www.linkedin.com/in/jinghezhang</a> <a href="https://github.com/JingheZ">https://github.com/JingheZ</a> <a href="http://people.virginia.edu/~jz4kg/">http://people.virginia.edu/~jz4kg/</a> Phone: 607-651-8906 Email: jz4kg@virginia.edu	
RESEARCH INTERESTS	<b>Machine Learning</b> , e.g., representation learning, metric learning; <b>Machine Learning Applications</b> , e.g., data mining, predictive modeling, recommender systems, health and medical informatics	
EDUCATION	<b>University of Virginia</b> Ph.D., Systems & Information Engineering GPA: 3.91/4.0	<i>08/2013 - Present</i> <i>Charlottesville, VA</i>
	<b>Binghamton University, State University of New York</b> Master of Science, Industrial & Systems Engineering GPA: 3.76/4.0	<i>08/2011 - 05/2013</i> <i>Binghamton, NY</i>
	<b>Hebei University of Technology</b> Bachelor of Science, Industrial Engineering GPA: 3.65/4.0	<i>09/2007 - 06/2011</i> <i>Tianjin, China</i>
ACADEMIC POSITIONS	<b>Graduate Research Assistant</b> Department of Systems & Information Engineering, University of Virginia	<i>08/2013 - Present</i>
	<b>Graduate Teaching Assistant</b> SYS6016 Machine Learning, University of Virginia	<i>01/2014 - 05/2014</i>
	<b>Graduate Research Associate</b> Watson Institute for Systems Excellence, Binghamton University Continuous Process Improvement in Healthcare, Quality Management Department, United Health Services (UHS)	<i>08/2011 - 05/2013</i>
PROJECT EXPERIENCE	<b>High-risk Patients Retrieval using Electronic Health Data</b> <ul style="list-style-type: none"><li>Performed representation learning on patients' medical histories to construct feature vectors with reduced information loss than using common methods</li><li>Conducted metric learning on the representation of patients' medical data and discovered informative features for clinical use</li><li>Trained 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</li><li>Evaluated the proposed framework by the early detection of anxiety/depression on new patients and achieved a 1%–4.5% higher accuracy compared to baseline methods</li></ul>	<i>09/2014 - Present</i>
	<b>Predictive Modeling on Hyperlactatemia Sepsis</b>	<i>05/2015 - Present</i>

- Analyzed 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.
- Examined different definitions of severe sepsis and its correlation with ICU, hospital, and 30-day mortalities
- To 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)
- To predict the real-time probability of lactate clearance considering a patient's demographic and physiological characteristics and clinical interventions

#### **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
- Implemented the above models using Spark on Hadoop clusters with significantly decreased computation time

#### **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, Naive 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.

#### **Spam Filtering Using Machine Learning Approaches** *09/2013 - 11/2013*

- Constructed static models using generalized linear regression based on principle component analysis and log transformation on the explanatory variables
- Modeled the trend, seasonality, and random fluctuation for spam and ham using time series data
- Combined the static and time series models using Bayes rule to further improve the performance of the spam filter

#### **Web Crawler and Document Analysis & Retrieval** *09/2014 - 11/2014*

- Programmed a web crawler for a medical forum in Java and extracted all posts in every threaded discussion about a particular topic, such as Diet, Hypertension, etc.
- Generated a document analyzer to tokenize, normalize, and stem the crawled documents, identified the most frequent words to compare with the standard stop words, and validated the Zipf's law

- Built an information retrieval system to retrieve similar documents of the input queries based on vector space model and popular language models, such as BM25

## Predictive Modeling on Hospital Readmission

04/2012 - 05/2013

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

## 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

## SKILLS

**Programming:** Python, Java, R, Matlab, SQL, HTML, CSS, JavaScript, Hadoop, Spark  
**Operating Systems:** Windows, Unix/Linux  
**Tools:** Latex, Git, SVN, Vim, Emacs, Arena, Simio, ExpertFit, Minitab, NetLogo  
**Certificates:** Lean Six Sigma Green and Black Belt from Dartmouth College  
**Languages:** English and Chinese