

## David C. Kale

*Ph.D. Candidate*

USC Information Sciences Institute, Marina del Rey, CA

dkale@usc.edu

<http://www-scf.usc.edu/~dkale/>

Github: <http://github.com/turambar>

Linkedin: <http://www.linkedin.com/in/davekale>

davekale@cs.stanford.edu

479-769-0570

### Research Interests and Overview

---

My research uses machine learning to extract insights from digital data in high impact domains, including health-care. Currently, I am studying the unique challenges that arise when applying deep learning to electronic health records. Because labeling by clinical experts is expensive, I am developing solutions that instead leverage domain knowledge to guide otherwise data-driven models. In previous work, I studied active learning and transfer learning, two complementary paradigms for supervised learning in the absence of carefully labeled data.

### Education

---

- 2012- ★ **Ph.D., Computer Science – University of Southern California**  
Advisor: Greg Ver Steeg. Topic: Machine Learning. *Expected graduation: May 2017*  
*Viterbi Dean's Doctoral Fellow and Alfred E. Mann Innovation in Engineering Doctoral Fellow*
- 2011 ★ **M.S., Computer Science – Stanford University**  
Specialization: Artificial Intelligence
- 2005 ★ **B.S. with distinction, Symbolic Systems – Stanford University**  
Inducted into *Phi Beta Kappa Academic Honor Society*.

### Professional and Research Experience

---

- 2015- ★ **Research Assistant – USC Information Sciences Institute**  
Study challenges that arise from the application of deep learning to clinical data, under the supervision of Profs. Greg Ver Steeg and Aram Galstyan. Also collaborate with Prof. Nigam Shah at the Stanford Biomedical Informatics Research Center.
- 2014 ★ **Research Intern – Microsoft Research, Redmond**  
Worked with the MSR medical devices team to develop novel wearable sensors for continuous monitoring of cardiovascular health.
- 2012–2015 ★ **Research Assistant – University of Southern California**  
Worked with Prof. Yan Liu on the application of machine learning to time series and clinical data, as well as active learning and transfer learning.
- 2011- ★ **Co-founder and Advisor – Podimetrics**  
Boston-area medical device start-up building a sensor and data analytics platform for detection and prevention of foot ulcers in patients suffering from diabetic neuropathy.
- 2009–2012 ★ **Data Scientist – Whittier VPICU, Children's Hospital Los Angeles**  
Applied statistical learning techniques to digital clinical data sets.  
Published manuscripts describing research results.

- 2006–2009      ★ **Research Assistant – Stanford Artificial Intelligence Lab**  
Worked on projects in computer vision, robotics, and automated driving.
- 2005            ★ **Engineering Intern – Google, Inc.**  
Extended web search framework to handle structured data for Froogle and Google Base.
- 2002            ★ **Research Assistant – SRI International**  
Implemented testing framework for Babylon speech-to-speech translation project.

## Research and Publications

---

### *In preparation or under review*

- H. Harutyunyan, H. Khachatrian, D. Kale, and Aram Galstyan. Multitask Learning and Benchmarking with Clinical Time Series Data. *Under review*. Available on arXiv.
- S. Dubois, N. Romano, D. Kale, N. Shah, K. Jung. Learning Effective Representations from Clinical Notes. *Under review*. Available on arXiv.

### *Refereed computer science*

- T. Quisel, L. Foschini, A. Signorini, and D. Kale. Collecting and Analyzing Millions of mHealth Data Streams. Proceedings of the 23rd ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2017.
- Z. Lipton,\* D. Kale,\* and R. Wetzel. Directly Modeling Missing Data in Sequences with RNNs: Improved Classification of Clinical Time Series. Proceedings of the Machine Learning for Healthcare Conference (MLCH), 2016.
- Z. Lipton,\* D. Kale,\* C. Elkan, and R. Wetzel. Learning to Diagnose with LSTM Recurrent Neural Networks. International Conference on Learning Representations (ICLR), 2016.
- Z. Che,\* D. Kale,\* W. Li, M.T. Bahadori, and Y. Liu. Deep Computational Phenotyping. Proceedings of the 21st ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2015.
- M.T. Bahadori, D. Kale, Y. Fan, and Y. Liu. Functional Subspace Clustering with Application to Time Series. Proceedings of the 31st International Conference on Machine Learning (ICML), 2015.
- D. Kale, M. Ghazvininejad, A. Ramakrishna, J. He, and Y. Liu. Hierarchical Active Transfer Learning. Proceedings of the SIAM International Conference on Data Mining (SDM), 2015.
- D. Kale,\* D. Gong,\* Z. Che,\* G. Medioni, R. Wetzel, P. Ross, and Y. Liu. An Examination of Multivariate Time Series Hashing with Applications to Health Care.. Proceedings of the IEEE 14th International Conference on Data Mining (ICDM), 2014.
- D. Kale and Y. Liu. Accelerating Active Learning with Transfer Learning. Proceedings of the IEEE 13th International Conference on Data Mining (ICDM), 2013.
- D. Kale and D. Stork. Estimating the Position of Illuminants in Paintings Under Weak Model Assumptions: An Application to the Works of Two Baroque Masters. In B. E. Rogowitz and T. N. Pappas (eds.), *Electronic Imaging: Human Vision and Electronic Imaging XIV*, vol. 7240, pp. 72401M112. SPIE/IS&T, Bellingham, 2009.

### *Refereed health and medical*

- D. Kale, Z. Che, M.T. Bahadori, W. Li, Y. Liu, and R. Wetzel. Causal Phenotype Discovery via Deep Networks. Proceedings of the American Medical Informatics Association (AMIA) Annual Symposium, 2015.

- D. Epstein, M. Reibel, J. B. Unger, M. Cockburn, L. A. Escobedo, D. Kale, J. C. Chang, and J. I. Gold. *The Effect of Neighborhood and Individual Characteristics on Pediatric Critical Illness*. *Journal of Community Health*, Feb. 2014.
- R. G. Khemani, E. B. Celikkaya, C. R. Shelton, D. Kale, P. A. Ross, R. C. Wetzel, and C. J. L. Newth. *Algorithms to estimate PaCO<sub>2</sub> and pH using non invasive parameters for children with Hypoxemic Respiratory Failure*. *Respiratory Care*, Dec. 2013.
- E. Ingram, D. Kale, and R. Balfour. *Hemilaminectomy for thoracolumbar Hansen Type I intervertebral disk disease in ambulatory dogs with or without neurologic deficits*. *Journal of Veterinary Surgery*: vol. 42, #8, pg. 924-931, Nov. 2013.
- B. Stubbs and D. Kale. *Sim\*TwentyFive: An Interactive Visualization System for Data-Driven Decision Support*. *Proceedings of the American Medical Informatics Association (AMIA) Annual Symposium*, 2012.
- B. Marlin, D. Kale, R. Khemani, and R. Wetzel. *Unsupervised Pattern Discovery in Electronic Health Care Data Using Probabilistic Clustering Models*. *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium (IHI)*, 2012.
- D. Crichton, C. Mattmann, A. Hart, D. Kale, R. Khemani, P. Ross, S. Rubin, P. Veeravatanayothin, A. Braverman, C. Goodale, and R. Wetzel. *An Informatics Architecture for the Virtual Pediatric Intensive Care Unit*. *Proceedings of the 24th IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, 2011.
- D. Kale, A. Hart, C. Mattmann, R. Khemani, P. Ross, P. Vee, J. Terry, R. Wetzel, and D. Crichton. *An Open Source, Grid-based Software Framework for Management and Sharing of Pediatric ICU Data*. *Proceedings of the 9th International Conference on Complexity in Acute Illness (ICCAI)*, 2010.

## Workshop

- S. Dubois, N. Romano, K. Jung, and N. Shah. *The Effectiveness of Transfer Learning in Electronic Health Records Data*. *International Conference on Learning Representations Workshop Track*, 2017.
- T. Quisel, D. Kale, and L. Foschini. *Intra-day Activity Better Predicts Chronic Conditions*. *NIPS 2016 Workshop on Machine Learning for Healthcare (NIPS ML4HC)*, 2016.
- K. Reing, D. Kale, G. Ver Steeg, and A. Galstyan. *Toward Interpretable Topic Discovery via Anchored Correlation Explanation*. *ICML Workshop on Human Interpretability in Machine Learning (WHI)*, 2016.
- Z. Lipton,\* D. Kale,\* and R. Wetzel. *Phenotyping of Clinical Time Series with LSTM Recurrent Neural Networks*. *NIPS 2015 Workshop on Machine Learning in Healthcare*.
- D. Kale, Z. Che, Y. Liu, and R. Wetzel. *Computational discovery of physiomes in critically ill children using deep learning*. *1st Workshop on Data Mining for Medical Informatics: Electronic Phenotyping*, *American Medical Informatics Association*, 2014.
- E. B. Celikkaya, C. Shelton, D. Kale, R. Wetzel, and R. Khemani. *Non-invasive Blood Gas Estimation for Pediatric Mechanical Ventilation*. *Proceedings of the NIPS 2013 Workshop on Machine Learning for Clinical Data Analysis and Healthcare*.
- D. Kale, S. Di, Y. Liu, and Y. Gil. *Capturing Data Analytics Expertise with Visualization in Workflows*. *AAAI Fall Symposium Series Discovery Informatics Workshop (DIS)*, 2013.
- D. Kale. *Unsupervised Pattern Discovery in Sparsely Sampled Clinical Time Series*. *At From Data to Knowledge Workshop*, *University of California Berkeley*, 2012.

## Other

- D. Kale and J. Patterson. *Conditional recurrent neural nets, generative AI Twitter bots, and DL4J*. Strata+Hadoop World, New York, 2016.
- J. Patterson, D. Kale, and Z. Lipton. *Deep learning and recurrent neural networks applied to electronic health records*. Strata+Hadoop World, San Jose, 2016.
- B. Sankaran, D. Kale, X. He, L. Cohen, and M. Ghazvininejad. *Learning and Optimization with Submodular Functions*. Available on arXiv.
- D. Kale. *Probabilistic Modeling of Electronic Health Records Data*. AI/ML Weekly Seminar at University of California, Irvine, Center for Machine Learning and Intelligent Systems. October 2011.
- D. Kale, B. Marlin, R. Khemani, and R. Wetzel. *A Novel Application of Unsupervised Learning to Electronic Health Care Records Data*. At *1st Southern California Workshop on Machine Learning (SoCaML 2011)*, 2011.
- D. Kale, B. Marlin, R. Khemani, and R. Wetzel. *Using Probabilistic Clustering to Find Patterns in Digital Medical Data*. At *1st Meaningful Use of Complex Medical Data Symposium (MUCMD 2011)*, 2011.
- A. Hart, D. Kale, R. Khemani, and H. Kincaid. *Distributed, Modular Grid Software for Data Management and Exploration of Patient-Centric Healthcare IT Information*. In proceedings of the O'Reilly Open Source Convention: Special Session on Healthcare Technology (OSCON 2010), 2010.

## Funded Grants

---

- 2009                   ★ **Advanced Computational Framework for Decision Support in Critically Ill Children**  
Co-PI, \$1,000,000 to Children's Hospital Los Angeles (NASA JPL sub-contractor)

## Teaching Experience

---

- Teaching Assistant   USC           CSCI 109: Introduction to Computing  
                          Stanford   CS 221: Introduction to Artificial Intelligence  
  CS 121: Introduction to Artificial Intelligence  
  CS 103A: Discrete Mathematics for Computer Science  
  CS 103B: Discrete Structures
- Section Leader       Stanford   CS 107: Computer Organization and Systems

## Professional Activities and Awards

---

- 2017                   ★ Judge – IBM Watson A.I. XPRIZE
- 2016                   ★ Organizer – NIPS 2017 Machine Learning for Health Care Workshop (MLH4)
- 2017                   ★ Organizer – Machine Learning for Health Care Conference (MLHC) 2017
- 2016                   ★ Organizer – NIPS 2016 Machine Learning for Health Care Workshop (MLH4)
- 2016                   ★ Organizer – Machine Learning for Health Care Conference (MLHC) 2016
- 2015                   ★ NSF Travel Award – 21st ACM Int'l Conf. on Knowl. Disc. and Data Mining (SIGKDD)
- 2014–2017           ★ Judge – Qualcomm Tricorder XPRIZE
- 2014                   ★ Student Travel Award – IEEE 14th Int'l Conference on Data Mining (ICDM)
- 2014–2017           ★ Alfred E. Mann *Innovation in Engineering* Fellowship  
                          One of three fellows conducting innovative interdisciplinary research.
- 2014                   ★ Organizer – AAAI 1st Modern AI for Health Analytics Workshop
- 2014                   ★ Program Committee – AAAI Discovery Informatics Workshop
- 2013                   ★ NSF Travel Award – IEEE 13th Int'l Conference on Data Mining (ICDM)
- 2013                   ★ Session Moderator – Health 2.0 7th Annual Fall Conference
- 2012–2013           ★ USC Viterbi School of Engineering Dean's Doctoral Fellowship  
                          Fellowship given to promising incoming Ph.D. students in engineering at USC
- 2011                   ★ Co-founder – Meaningful Use of Complex Medical Data Symposium (MUCMD)
- 2010–2013           ★ Organizer – LA Machine Learning Meetup

- 2011–  
★ **Co-founder and advisor, *Podimetrics***  
★ **Winner of Life Sciences Track – MIT100K Business Plan Competition**  
★ **Second Runner-Up – MIT100K Elevator Pitch Competition**  
★ **Winner – MIT Hacking Medicine Pitching Competition**  
Co-founded medical device company *Podimetrics*, which is building a remote monitoring platform for early detection of diabetic foot ulcers; raised A round funding in early October of 2012.
- 2010–  
★ **Project Management Committee – OODT Project, Apache Software Foundation**
- 2003  
★ **Inducted into Phi Beta Kappa Academic Honor Society**  
Stanford University, Beta Chapter of California; inducted as junior (top 5% of class).

## U.S. Patents

---

- 2012-2013  
★ **Patents pending – Podimetrics**  
Podimetrics has filed several provisional patents related to detection and prediction of diabetic foot ulcers.
- 2011  
★ **Patent #7,933,900 (granted April 26, 2011; filed October 23, 2005) – Google, Inc.**  
*Abstract:* A user can refine a search over structured data by specifying that a label or an attribute value be used to further filter the results of a query.  
*Inventors:* Bindu Reddy, Jonathan Brunsman, Ning Mosberger, Gaurav Bhaya, Sarah Sirajuddin, David Kale, Jennifer Kozenski, Arvind Sundararajan, and Puneet Agarwal.