

# Balaji Krishnapuram

1021 Penn Circle, Apt E-612  
King of Prussia, PA-19406  
Fax: (610)-448-4274

FirstName dot LastName at GMail dot com  
Phone: (484)-831-5478 (H)  
Phone: (301)-312-2150 (C)

---

## Research Interests

Machine Learning, Artificial Intelligence, and Statistical Data Mining. Application areas: medical decision support systems, information retrieval, web search, computational advertising and computational finance.

## Education

- ◆ Ph. D. in Electrical & Computer Engineering, Duke University Sept '04  
**Dissertation:** Adaptive classifier design using labeled and unlabeled data  
**Contributions:** Developed algorithms for
  - Supervised, semi-supervised and unsupervised statistical pattern recognition; Active learning
  - Applications to: problems in cancer diagnosis, detection of land-mines & under-water mines
- ◆ MS in Electrical & Computer Engineering, Duke University June '01  
**Thesis:** Multi-aspect target detection in SAR imagery
- ◆ B.Tech. in Electrical Engineering, Indian Institute of Technology (IIT), Kharagpur, India June '99  
**Dissertation:** Wavelet Vector-Quantization based still grayscale Image Compression

## Honors and Awards

- ◆ Graduate Research Assistantship, Dept of Electrical Engineering, Duke University Aug '99-Sep '04
- ◆ Travel-grant for attending international research conference, RECOMB Apr '03
- ◆ Recipient of IMO Math Scholarship awarded by the Dept. of Atomic Energy, Govt. of India '95-'99  
Ranked 27<sup>th</sup> in the Indian National Math Olympiad (INMO), 1994
- ◆ Recipient of National Science Talent Scholarship (NTSE) awarded by Govt. of India '94-'99

## Work Experience

### Current (full-time) position

- ◆ Senior Staff Scientist, Dec '07-present  
Staff Scientist, Nov '04-Nov '07  
Siemens Medical Solutions USA, Image & Knowledge Management, CAD group (Malvern, PA)  
At Siemens Medical Solutions, I have developed novel products in three broad areas. First, I have developed several *computer aided diagnosis* (CAD) products that automatically identify early stage cancer of the Breast, Lung, and Colon based on X-ray, and CT images. These systems ensure that malignant cancers are not accidentally overlooked by radiologists.

Second, I have worked on *personalized therapy-selection*. Modern medicine has placed increasing emphasis on choosing optimal therapy for groups of patients (eg T1 stage Cancer patients) based on the population averages of outcomes from clinical trials. However, this approach ignores the large heterogeneity among the patients in any such group, and also does not explicitly account for the extent of medical information that would be missing for any patient. The systems I developed are personalized to predict the optimal therapy for each patient, based on the available data.

Third, I have developed *medical decision support systems* and medical data mining systems that combine all available structured and unstructured data for a patient, including textual notes from doctors, medical images, billing records, lab records, etc. Such systems support retrospective analysis of a hospital's compliance with various quality measures. These systems can also automatically identify patients who are eligible for participation in various clinical trials.

## Internships

- ◆ Research-Intern, Microsoft Research Ltd. (Cambridge, UK) summer '03  
I developed a probabilistic model for ink in user drawn figures on Tablet PCs. Based on this model, I designed algorithms for understanding hand-drawn sketches as a new user-interface for MS-Office.
- ◆ Intern, ITC Ltd., the Indian subsidiary of BAT plc. (India) summer '98  
I developed an expert-system for quality control. The implementation of this system resulted in an 8% improvement in internal quality metrics.

## Invited Book Chapters

1. Balaji Krishnapuram, Lawrence Carin, and Alexander Hartemink, "Gene expression analysis: Joint feature selection and classifier design," in *Kernel Methods in Computational Biology*, B. Scholkopf, K. Tsuda, and J.-P. Vert (editors), pp. 299-318, MIT press, 2004.
2. Ya Xue, Xuejun Liao, Lawrence Carin, and Balaji Krishnapuram, "Multi-Task learning for classification with Dirichlet process priors," accepted for publication in *Inductive Transfer*, D. Silver, K. Bennet, R. Caruana (editors), Springer Verlag, 2007.

## Refereed Journal Publications

3. Balaji Krishnapuram, Jefferey Sichina, and Lawrence Carin, "Physics based detection of targets in SAR imagery using support vector machines," *IEEE Sensors Journal*, Vol. 3, No. 2, pp. 147-158, April 2003.
4. Balaji Krishnapuram, Lawrence Carin, and Alexander Hartemink, "Joint classifier and feature optimization for comprehensive cancer diagnosis using gene expression data," *Journal of Computational Biology*, Vol. 11, pp 227--242, March 2004.
5. Balaji Krishnapuram, Alexander Hartemink, Lawrence Carin, and Mario Figueiredo, "A Bayesian approach to joint feature selection and classifier design," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, pp. 1105-1111, September 2004.
6. Yijun Yu, Balaji Krishnapuram, and Lawrence Carin, "Inverse scattering with sparse Bayesian vector regression," *Inverse Problems*, Special Issue on Electromagnetic Characterization of Buried Obstacles, Vol. 20, No. 6, pp. S217-S231, December 2004.
7. Balaji Krishnapuram, Lawrence Carin, Mario Figueiredo, and Alexander Hartemink, "Sparse multinomial logistic regression: fast algorithms, and generalization bounds," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 27, No. 6, pp. 957-968, June 2005.
8. Shihao Ji, Balaji Krishnapuram, and Lawrence Carin, "Variational Bayes for continuous hidden Markov models and its application to active learning," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 28, No 4, pp. 522-532, April 2006.
9. Ya Xue, Xuejun Liao, Lawrence Carin, and Balaji Krishnapuram, "Multi-Task learning for classification with Dirichlet process priors," *Journal of Machine Learning Research*, Vol 8, pp 33-63, Jan 2007.
10. David Williams, Ya Xue, Xuejun Liao, Lawrence Carin, and Balaji Krishnapuram, "On incomplete data classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 29, No. 3, pp. 427-436, Mar 2007.
11. R. Seigneuric, M.H.W. Starmans, G. Fung, Balaji Krishnapuram, D.S.A. Nuyten, A. van Erk, M.G. Magagnin, K.M. Rouschop, S. Krishnan, R. Bharat Rao, C.T.A. Evelo, A.C. Begg, B.G. Wouters, P. Lambin, "Impact of a supervised gene signature of early hypoxia on patient survival", accepted to *Radiotherapy & Oncology*, 2007.
12. Vikas Raykar, Ramani Duraiswami, Balaji Krishnapuram, "Efficient algorithms for learning preference relations," accepted to *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2007.
13. Glenn Fung, Murat Dundar, Balaji Krishnapuram, and R. Bharat Rao, "Multiple instance learning via alternate optimization," accepted to *IEEE Trans. Biomedical Engineering*, 2007.

## Refereed Conference Publications

14. Eric Jones, Jiangqi He, Balaji Krishnapuram, John Pormann, John A. Board, and Lawrence Carin, "An electromagnetic simulation and SAR processing environment," 2001 SPIE AeroSense Conference, Proceedings of SPIE, Vol. 4367, Orlando, FA, April 2001.
15. Balaji Krishnapuram and Lawrence Carin, "Support vector machines for improved multi-aspect target recognition using the fisher kernel scores of hidden markov models," 2002 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Vol. 3, pp 2989-2992, IEEE Press, Orlando, FA, May 2002.
16. Balaji Krishnapuram, Lawrence Carin and Alexander J. Hartemink, "Applying logistic regression and RVM to achieve accurate probabilistic cancer diagnosis from gene expression profiles," 2002 IEEE Workshop on Genomic Signal Processing and Statistics (GENSIPS) , IEEE Press, Raleigh, NC, October 2002.
17. Balaji Krishnapuram, Lawrence Carin and Alexander J. Hartemink, "Joint classifier and feature optimization for cancer diagnosis using gene expression data," The Seventh Annual International Conference on Research in Computational Molecular Biology (RECOMB) 2003, ACM press, Berlin, Germany, April 2003.
18. Qiuhua Liu, Balaji Krishnapuram, Pallavi Pratapa, Xuejun Liao, Alexander Hartemink and Lawrence Carin, "Identification of differentially expressed proteins using MALDI-TOF mass spectra," 2003 Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, November 2003.
19. Xuejun Liao, Hui Li and Balaji Krishnapuram, "An M-ary KMP classifier for multi-aspect target

- classification," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Montreal, Canada, May 2004.
20. Balaji Krishnapuram, Christopher Bishop and Martin Szummer, "Generative Bayesian models for shape recognition," The Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR-9), Tokyo, Japan, October 2004.
  21. Balaji Krishnapuram, David Williams, Ya Xue, Lawrence Carin, Alexander Hartemink, and Mario Figueiredo, "On semi-supervised classification," Neural Information Processing Systems (NIPS), Vancouver, Canada, December 2004.
  22. Balaji Krishnapuram, David Williams, Ya Xue, Lawrence Carin, Mario Figueiredo, and Alexander Hartemink, "Active learning of features and labels," Workshop on learning with multiple views at the 22<sup>nd</sup> International Conference on Machine Learning (ICML-05), Bonn, Germany, August 2005
  23. Glenn Fung, Romer Rosales, and Balaji Krishnapuram, "Learning rankings via convex hull separation," Neural Information Processing Systems (NIPS), Vancouver, Canada, December 2005.
  24. Ya Xue, Xuejun Liao, Lawrence Carin, and Balaji Krishnapuram, "Learning multiple classifiers with Dirichlet process mixture priors," Workshop on Nonparametric Bayesian methods, Neural Information Processing Systems (NIPS), Vancouver, Canada, December 2005.
  25. Glenn Fung, Balaji Krishnapuram, Nicolas Merlet, Eli Ratner, Philippe Bamberger, Jonathan Stoeckel and R. Bharat Rao, "Addressing image variability while learning classifiers for detecting clusters of micro-calcifications," International Workshop on Digital Mammography (IWDM), Manchester, UK, June 2006
  26. Volkan Vural, Glenn Fung, Balaji Krishnapuram, and Jennifer Dy, "Batch Classification with applications to Computer Aided Diagnosis," European Conference on Machine Learning (ECML), Berlin, Germany, August 2006.
  27. Vivian van den Boogaart, Annemarie Dingemans, Victor Thijssen, Robert-Jan van Suylen, Balaji Krishnapuram, Arjan Griffioen, "Angiogenesis gene expression profiling as prognostic marker in non-small cell lung cancer", October 2006.
  28. Glenn Fung, Murat Dundar, Balaji Krishnapuram, and R. Bharat Rao, "Multiple Instance Algorithms for Computer Aided Diagnosis," Neural Information Processing Systems (NIPS), Vancouver, Canada, December 2006.
  29. Murat Dundar, Balaji Krishnapuram, Jinbo Bi, and R. Bharat Rao, "Learning classifiers when the training data is not IID," International Joint Conference on Artificial Intelligence (IJCAI) January 2007
  30. Vikas Raykar, Ramani Duraiswami, and Balaji Krishnapuram, "A fast algorithm for learning large scale preference relations," International Conference on Artificial Intelligence and Statistics (AISTATS), Puerto Rico, March 2007
  31. Balaji Krishnapuram, C. Dehing, H. Steck, H. van der Weide, D. De Ruyscher, B. Nijsten, S. Wanders, L. Boersma, R.B. Rao, and Ph. Lambin, "A knowledge-model for predicting radiation-induced Esophagitis," 49<sup>th</sup> Annual meeting of the American Society for Therapeutic Radiology and Oncology (ASTRO), Nov 2007.
  32. Shipeng Yu, Balaji Krishnapuram, Romer Rosales, Harald Steck, and R. Bharat Rao, "Bayesian Co-training," Neural Information Processing Systems (NIPS), Vancouver, Canada, December 2007.
  33. Vikas Raykar, Harald Steck, Balaji Krishnapuram, Cary Dehing-Oberije, and Philippe Lambin, "On ranking in survival analysis: bounds on the concordance index," Neural Information Processing Systems (NIPS), Vancouver, Canada, December 2007.
  34. Vikas Raykar, Harald Steck, Balaji Krishnapuram, Cary Dehing-Oberije, and Philippe Lambin, "On ranking in survival analysis: bounds on the concordance index," Neural Information Processing Systems (NIPS), Vancouver, Canada, December 2007.
  35. Balaji Krishnapuram, Jonathan Stoeckel, Vikas Raykar, R. Bharat Rao, Philippe Bamberger, Eli Ratner, Nicolas Merlet, Inna stainvas, Menahem Abramov, and Alexandra Manevitch, "Multiple instance learning improves CAD detection of masses in digital mammography," International Workshop on Digital Mammography (IWDM), Tucson, AZ, June 2008.
  36. Isaac Leichter, Richard Lederman, Eli Ratner, Nicolas Merlet, Glenn Fung, Balaji Krishnapuram, and Philippe Bamberger, "Does a mammography CAD algorithm with varying filtering levels of detection marks, used to reduce the false mark rate, adversely affect the detection of small masses?," International Workshop on Digital Mammography (IWDM), Tucson, AZ, June 2008.
  37. Vikas Raykar, Balaji Krishnapuram, Murat Dundar, Jinbo Bi, and R. Bharat Rao "Bayesian multiple instance learning: automatic feature selection and inductive transfer," 25<sup>th</sup> International Conference on Machine Learning (ICML), Helsinki, Finland, July 2008.

### **Submitted to Refereed Journals (Under Review)**

38. Volkan Vural, Glenn Fung, Balaji Krishnapuram, and Jennifer Dy, "Batch-wise classification with applications to Computer Aided Diagnosis," submitted for review to IEEE Trans. Pattern Analysis and

Machine Intelligence.

39. Murat Dundar, Balaji Krishnapuram, Jinbo Bi, and R. Bharat Rao, "Learning classifiers from non IID data," submitted for review to Pattern Recognition.
40. M.H.W. Starmans, B. Krishnapuram, H. Steck, D.S.A. Nuyten, R. Seigneuric, F.M. Buffa, A.L. Harris, B.G. Wouters, P. Lambin, "Clinical relevance of a knowledge-based proliferation signature: the prognostic value in published patient microarray studies", submitted for review to the New England Journal of Medicine.
41. Shipeng Yu, Balaji Krishnapuram, Romer Rosales, Harald Steck, and R. Bharat Rao, "Bayesian Multi-view Learning," submitted for review to IEEE Trans. Pattern Analysis and Machine Intelligence.
42. Vikas Raykar, Balaji Krishnapuram, Murat Dundar, and R. Bharat Rao, "Probabilistic models for Multiple Instance Learning, and their application in active data acquisition", to be submitted to IEEE Trans. Pattern Analysis and Machine Intelligence.

### **Product/Technology Demonstrations at International Conferences**

43. R. Bharat Rao, Romer Rosales, Stefan Niculescu, Sriram Krishnan, Luca Bogoni, Xiang S. Zhou, and Balaji Krishnapuram, "Mining medical records for Computer Aided Diagnosis," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Philadelphia, PA, August 2006
44. Murat Dundar, Balaji Krishnapuram, Glenn Fung, R. Bharat Rao, "Early Stage Cancer Diagnosis," Neural Information Processing Systems (NIPS), Vancouver, Canada, Dec 2006.

### **Freely Distributed Research Software**

- ◆ Software implementation of the mathematical algorithms designed as part of my Ph. D. thesis '03-'04  
Code developed as part of my PhD thesis is used by researchers in over 50 research laboratories in 10 countries around the world. This work has been used by researchers in diverse fields such as: computer vision, speech recognition, financial data analysis, landmine-detection, cancer diagnosis, gene expression analysis in functional genomics, and proteomics. Papers describing the code have been cited over 100 times.

### **Invited Talks**

- ◆ "Modern Statistical Machine Learning tools for Applied Clinical Research" Sept '06  
Maastricht Radiation Oncology (MAASTRO) clinic, GROW Research Institute, Maastricht, Netherlands
- ◆ "Applications of Modern Machine Learning Technologies in Clinical Decision Support Systems" Sept '06  
Department of Machine Learning & Data Mining, Maastricht University, Maastricht, Netherlands
- ◆ "Classifying Non-IID data," Jul '06  
Workshop on Adaptive Sensing and Waveform Scheduling,  
Sponsors: Defense Advanced Research Project agency (DARPA) & Army Research Office (ARO)  
Duke University, Durham, NC
- ◆ "Bayesian Semi-parametric models for multi-task learning," Aug '05  
Workshop on Adaptive Multi-Modal sensing and Waveform Scheduling,  
Sponsors: Defense Advanced Research Project agency (DARPA) & Army Research Office (ARO)  
Georgia Institute of Technology, Atlanta, GA
- ◆ "Learning classifiers under a limited budget for acquiring training data," Jun '05  
2005 Joint Annual Meeting of the Interface and Classification Society of North America (CSNA),  
Washington University in St. Louis, MO
- ◆ "Autonomous learning of multi-sensor classifiers from labeled and unlabeled data," Jul '04  
2004 Meeting of the International Federation of Classification Societies (IFCS),  
Chicago, IL
- ◆ "Active and semi-supervised learning for classifier design," Nov '03  
Statistical and Applied Mathematical Sciences Institute (SAMSI), Theory & Methods group,  
Research Triangle Park, NC

### **Professional Service Activities**

- ◆ Member of Journal Editorial Board '06-present  
Open Electrical and Electronic Engineering Journal, Open Artificial Intelligence Journal
- ◆ Session chair in the 2004 meeting of the International Federation of Classification Societies Jul '04  
IFCS-2004, Chicago, IL

- ◆ Reviewer for International Research Journals: *'02-present*  
 Journal of Machine Learning Research, Machine Learning Journal, IEEE Trans. Pattern Analysis and Machine Intelligence, IEEE Trans. Knowledge and Data Engineering, Neurocomputing, IEEE Trans. Neural networks, IEEE Signal Processing Letters, IEEE Trans. Signal Processing, EURASIP Journal on Applied Signal Processing, IEEE Trans. Circuits and Systems (B), Bioinformatics, IEEE/ACM Trans. Computational Biology and Bioinformatics, Technometrics, Computer Methods & Programs in Medicine, Journal of Computer Science & Technology (Chinese Academy of Sciences)
- ◆ Reviewer for International Research Conferences (only highly respected confs. listed): *'04-present*  
 Pacific Symposium on Biocomputing (PSB), International Conference on Machine Learning (ICML), International Conference on Knowledge Discovery and Data Mining (SIGKDD), Neural Information Processing Systems (NIPS).

### **Other Service Activities**

- ◆ Teacher in Hindu Sunday School, Chinmaya Mission Balavihar program *'03-'04*
- ◆ Graduate student representative in the South Asian student organization at Duke University *'99-'00*
- ◆ Library secretary of dormitory with administrative & financial responsibilities *'98-'99*