Jennifer G. Dy

Assistant Professor
Department of Electrical and Computer Engineering
Northeastern University, Boston, MA 02115
(617) 373-3062
email: jdy@ece.neu.edu
http://www.ece.neu.edu/faculty/jdy/

RESEARCH INTERESTS

Machine Learning, Data Mining, Pattern Recognition, Medical Image Analysis.

EDUCATION

Ph.D. in Electrical and Computer Engineering, 2001, Purdue University, West Lafayette, IN
Machine Learning and Robot Vision Laboratories. Supported by NIH.
Thesis: *Feature Selection for Unsupervised Learning Applied to Content-Based Image Retrieval*. Supervisors: Carla Brodley and Avi Kak

Master of Science in Electrical and Computer Engineering, 1997, Purdue University
Electronic Imaging Systems Laboratory. Supported by Hewlett-Packard.
Project: *Bitmap Resolution Synthesis*. Supervisor: Jan Allebach

Bachelor of Science in Electrical Engineering, 1993,
University of the Philippines, Quezon City, Philippines

HONORS, AWARDS, AND SCHOLARSHIPS

NSF CAREER Award (2004)
B.S. Magna Cum Laude (1994)
Most Outstanding All-Around Graduating Student of the College of Eng. (1994)
Best Undergraduate Research Project of the College of Engineering (1994)
BPI Philippines National Science Award (1994)
University of the Philippines Presidential Scholarship, tuition scholarship (1991-1993)
Gerry Roxas Foundation Scholarship, tuition scholarship (1989-1990)

SUMMARY OF ACHIEVEMENTS

- NSF CAREER Award, 2004
- Total external funding (including NSF, Army, MGH, and BBN Technologies): $1,073,653.
- Total internal funding from CDSP and the RSDF program: $35,000.
- 1 Ph.D. student graduated. 3 MS students graduated. 6 Ph.D. students in progress.
- 20 publications from journals and highly selective conferences.
- Action Editor, *Machine Learning*. Program Committee Member for several conferences, including *ICML, KDD, AAAI, SDM*. Publications Chair, *ICML 04*.

APPOINTMENTS

2002-present	Assistant Professor, Northeastern University, Boston, MA
1996-2000	Research Assistant, Purdue University, West Lafayette, IN
1993-1995	Instructor, University of the Philippines

GRANTS

EXTERNAL GRANTS (Total: $1,073,653):

NSF CAREER: A Foundation for Unsupervised Learning of High-Dimensional Data PI
Start date: March 1, 2004 Estimated expiration date: February 28, 2009
Expected total amount: $507,394.

Army STTR Phase II: In-Building Acoustic Signature Identification and Localization, co-PI
Start date: August 31, 2006 Estimated expiration date: August 30, 2008
Expected total amount for Northeastern University: $263,443.

NSF Major Research Instrumentation: MRI: Enabling Research on Terabyte-Scale Datasets, co-PI
Start date: August 1, 2006 Estimated expiration: July 31, 2008
Expected total amount: $199,000.

BBN Technologies, PI
Start date: January 1, 2007 Estimated expiration: August 31, 2007
Expected total amount: $34,657.

Mass. General Hospital: Algorithms for Tracking Tumors in Fluoroscopic Images
July 1, 2006 – August 31, 2006 $4,078.

The Methodist Hospital Research Institute: Algorithms for Segmentation and Tracking of Neuron Images
July 1, 2007 – June 30, 2008 $26,058.

Center for Subsurface Sensing and Imaging Systems (CenSSIS), an NSF Engineering Research Center, Senior Investigator
January 2003 – October 2003, $35,528.

Xilinx, Inc. Equipment Donation
Jan. 1, 2004 – Dec. 31, 2004, $3,495.

INTERNAL GRANTS (Total: $35,000):

Research and Scholarship Development Fund (RSDF) award: Identifying Metabolic Bottlenecks in the Production of Valuable Anti-Cancer Compounds from Catharanthus roseus Cell Cultures through Proteomics and Data Mining, co-PI
July 2006 – December 2007 Total amount: $25,000.

Communications and Digital Signal Processing (CDSP) center seed grant: Automated Processing and Classification of Confocal Skin Images, PI
May 1, 2006 – August 2006 Expected total amount: $10,000.

BOOK CHAPTER

J. G. Dy, "Unsupervised Feature Selection," invited book chapter in **Computational Methods of Feature Selection**, edited by Huan Liu and Hiroshi Motoda, *Chapman and Hall/CRC Press*, to appear 2007.

JOURNAL PUBLICATIONS

1. V. Vural, G. Fung, J. G. Dy, and B. Rao, *"Fast Semi-supervised Classifiers Using A-priori Metric Information,"* **Optimization Methods and Software, Special Issue on Mathematical Programming in Machine Learning and Data Mining**, to appear.

2. Y. Cui, J. G. Dy, G. C. Sharp, B. Alexander, and S. B. Jiang, *"Multiple Template Based Fluoroscopic Tracking of Lung Tumor Mass without Implanted Fiducial Markers,"* **Physics in Medicine and Biology**, to appear.

3. F. Azmandian, D. Kaeli, J. Dy, E. Hutchinson, M. Ancukiewicz, A. Niemierko, and S. B. Jiang, *"Towards the Development of an Error Checker for Radiotherapy Treatment Plans: A Preliminary Study,"* **Physics in Medicine and Biology**, to appear.

4. T. Su and J. G. Dy, "*In Search of Deterministic Methods for Initializing K-Means and Gaussian Mixture Clustering,"* **Intelligent Data Analysis**, Vol. 11, No. 4, pp. 319-338, 2007.

5. Y. Cui, J. G. Dy, G. C. Sharp, B. Alexander and S. B. Jiang, *"Robust Fluoroscopic Respiratory Gating for Lung Cancer Radiotherapy without Implanted Fiducial Markers,"* **Physics in Medicine and Biology**, Vol. 52, 2007, pp. 741-755.

6. J. Dy, *"Feature Selection for Unlabeled Data,"* Evolving Feature Selection, **IEEE Intelligent Systems, Trends & Controversies**, Vol. 20, No. 6, pp. 66-68, November/December, 2005.

7. J. G. Dy and C. E. Brodley, "*Feature Selection for Unsupervised Learning,"* **Journal of Machine Learning Research**, Volume 5, pp. 845-889, August 2004.

8. J. G. Dy, C. E. Brodley, A. Kak, L. S. Broderick, and A. M. Aisen, *"Unsupervised Feature Selection Applied to Content-Based Retrieval of Lung Images,"* **IEEE Transactions on Pattern Analysis and Machine Intelligence**, Vol. 25, No. 3, pp. 373-378, March 2003.

9. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, A. Marchiori, *"Automated Storage and Retrieval of Medical Images to Assist Diagnosis: Implementation and Preliminary Assessment,"* **Radiology,** 228(1), pp. 265-270, July 2003.

HIGHLY SELECTIVE CONFERENCE PUBLICATIONS

10. Y. Cui, X. Fern, and J. G. Dy, "Non-redundant Multi-view Clustering Via Orthogonalization," *Proceedings of the IEEE International Conference on Data Mining*, Omaha, NE, October 2007. (Acceptance rate: 19.2%, among the 7.2% accepted as regular papers.)

11. V. Vural, G. Fung, B. Krishnapuram, J. Dy, and B. Rao, "Batch Classification with Applications in Computer Aided Diagnosis," *Proceedings of the Seventeenth European Conference on Machine Learning*, vol. 4212, p. 449-460, Berlin, Germany, Sept. 18-22, 2006. (Acceptance rate: 25.5%, among the 14.5% accepted as full papers.)

12. K. Sanghai, T. Su, J. G. Dy, and D. Kaeli, "A Multinomial Clustering Model for Fast Simulation of Computer Architecture Designs," *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 808-813, August 21-24, Chicago, IL, 2005. (Acceptance rate: 34.2%).

13. B. Cordes, J. Dy, M. Leeser, and J. Goebel, "Enabling a Real-Time Solution for Neuron Detection with Reconfigurable Hardware," *IEEE International Workshop on Rapid System* Prototyping, pp. 128-134, 2005. (Acceptance rate: 53.8%).

14. T. Su and J. G. Dy, "A Deterministic Method for Initializing K-means Clustering," *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence*, pp. 784-786, November, 2004, Boca Raton, Florida. (Acceptance rate: 49.3%).

15. T. Su and J. G. Dy, "Automated Hierarchical Mixtures of Probabilistic Principal Component Analyzers," *Proceedings of the 21st International Conference on Machine Learning*, pp. 775-782, Banff, Alberta, Canada, July 2004. (Acceptance rate: 32%, among the 17.7% unconditionally accept papers.)

16. V. Vural and J. G. Dy, "A Hierarchical Method for Multi-Class Support Vector Machines," *Proceedings of the 21st International Conference on Machine Learning*, pp. 831-838, Banff, Alberta, Canada, July 2004. (Acceptance rate: 32%, among the 17.7% unconditionally accept papers.)

17. J. G. Dy and C. E. Brodley, "Visualization and Interactive Feature Selection for Unsupervised Data," *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 360-364, August 20-23, 2000, Boston, MA. (Acceptance rate: 20%).

18. J. G. Dy and C. E. Brodley, "Feature Subset Selection and Order Identification for Unsupervised Learning," *Proceedings of the Seventeenth International Conference on Machine Learning*, pages 247-254, June 29-July 2, 2000, Stanford University, CA. (Acceptance rate: 43%, among the 19% unconditionally accept papers.)

19. C. E. Brodley, A. C. Kak, J. G. Dy, C. R. Shyu, A. Aisen, and L. Broderick, "Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision," *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, pp. 760-767, July 18-22, 1999, Orlando, FL. (Acceptance rate: 27%).

20. J. G. Dy, C. E. Brodley, A. Kak, C. Shyu and L. S. Broderick, "The customized-queries approach to CBIR using EM," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Vol. II, pages 400-406, June 1999, Fort Collins, Colorado. (Acceptance rate: 38%).

OTHER PEER REVIEWED CONFERENCES

J. Fan, X. Zhou, J. Dy, Y. Zhang and S. Wong, "Spine Detection Using Curvilinear Structure Detector and Spine Tracking in In Vivo Image," *Third IEEE-NIH Life Science Systems and Application (LISA) Workshop*, November 2007, to appear. (*Paper*)

J. G. Dy, C. E. Brodley, A. Kak, C. Shyu and L. S. Broderick, "The customized-queries approach to CBIR," *Proceedings of the Storage and Retrieval for Image and Video Databases VII, IS&T/SPIE Electronic Imaging '99*, Vol. 3656, pages 22-32, January 1999, San Jose, CA. (*Paper*)

J. G. Dy and J. Allebach, "Bitmap Resolution Synthesis," *1997 IEEE Workshop on Nonlinear Signal and Image Processing*, September 8-10, 1997, Grand Hotel on Mackinac Island, Michigan. (*Paper*)

Y. Cui, J. G. Dy, G. C. Sharp, B. Alexander, and S. B. Jiang, "Fluoroscopic Tracking of Lung Tumor Mass without Implanted Fiducial Markers," *American Association of Physicists in Medicine (AAPM) 48th Annual Meeting*, *Medical Physics*, Vol. 33, Issue 6, p. 2162, Orlando, FL, July 30-August 3, 2006. (*Abstract*)

Y. Cui, J. G. Dy, G. C. Sharp, B. Alexander, and S. B. Jiang, "Correlation Score Based Respiratory Gating for Lung Cancer Radiotherapy without Implanted Fiducial Markers," *American Association of Physicists in Medicine (AAPM) 48th Annual Meeting*, *Medical Physics*, Vol. 33, Issue 6, p. 2244, Orlando, FL, July 30-August 3, 2006. (*Abstract*)

PROFESSIONAL ACTIVITIES

Action Editor, *Machine Learning (2007-present)*
Editorial Board Member, *Machine Learning (2004-2007)*
Publications Chair, *International Conference on Machine Learning (2004)*
Senior Program Committee Member, *International Conference on Machine Learning (2007)*
Program Committee Member,

- *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2001, 2003, 2005, 2006, and 2007),*
- *International Conference on Machine Learning (2005 and 2006),*
- *National Conference on Artificial Intelligence (AAAI 2005 and 2007),*
- *SIAM International Conference on Data Mining (2004),*
- *SIAM Data Mining (SDM) Workshop on Feature Selection (2005),*
- *SIAM Data Mining (SDM) Workshop on Clustering High-Dimensional Data (2005).*

Reviewer,
Machine Learning,
Journal of Machine Learning Research,
Journal of Artificial Intelligence Research,
IEEE Transactions on Pattern Analysis and Machine Intelligence,
Pattern Recognition Letters,
Knowledge and Information Systems,
Bioinformatics,
Encyclopedia of Computer Science and Engineering,
Data Mining and Knowledge Discovery,
Information Fusion
IEEE Transactions on Geoscience and Remote Sensing
National Science Foundation (2003, 2004, 2005)
External Reviewer,
International Conference on Machine Learning,
IEEE Conference on Computer Vision and Pattern Recognition,
IEEE Workshop of Content-Based Access of Image and Video Databases
Member, IEEE, ACM

INVITED TALKS

- "Clustering High-Dimensional Data," **A Workshop for Women in Machine Learning**, San Diego, CA, October 4, 2006.
- **"**Clustering High-Dimensional Data," **BBN Technologies**, Cambridge, MA, Oct. 27, 2006.
- "Clustering High-Dimensional Data," Joint Seminar between the Department of Electrical and Computer Engineering and the Department of Computer Science, **Tufts University**, April 18, 2006.
- "Clustering High-Dimensional Data and Multi-Class Support Vector Machines," **Motorola Labs**, August 25, 2005.
- "Clustering High-Dimensional Data," Department of Computer Science, **Boston University**, November 3, 2004.

- "Introduction to Support Vector Machines," Machine Learning Class, Computer Science Department, **Tufts University**, September 30, 2004.
- "Feature Selection for Unsupervised Learning Applied to Content-Based Image Retrieval," Statistical AI Reading Group, **Massachusetts Institute of Technology**, June 27,2002.
- "Feature Selection for Unsupervised Learning Applied to Content-Based Image Retrieval," Information and Computer Science, **University of California, Irvine**, January 9, 2001.
- "Feature Selection for Unsupervised Learning", Department of Computer Science, **Oregon State University**, November 20, 2000.
- "Feature Selection and Interactive Visualization for Unsupervised Learning", **Silicon Graphics**, October 6, 2000.

INVITED PANEL

- "Panel topic: What can I do with a Ph.D.? Perspectives on research, teaching, and industry careers," **Young Professionals Panel, AAAI(American Association for Artificial Intelligence)/SIGART(ACM Special Interest Group on Artificial Intelligence) Doctoral Consortium 2006**, July 17, 2006.

GRADUATE STUDENTS:

Current Ph.D. Students:

Volkan Vural, Ying Cui, Sila Kurugol, Yan Yan, Yue Guan, Jing Fan

Ph. D. Graduates:

1. Ting Su (April 2007).
   Thesis: *Clustering High-Dimensional Data*
   She is currently at Mathworks.

M.S. Graduates:

1. Ting Su (May 2003).
   Project: *Non-Random Initialization Method for K-Means Clustering*
2. Volkan Vural (April 2004).
   Project: *A Hierarchical Method for Multi-Class Support Vector Machines*
3. Hongyan Liu (April 2005).
   Project: *A Comparative Study of Feature Subspace Selection Algorithms for Unsupervised Learning*

Undergraduate Project Advising:

1. Sufeng Li (April 2005).
   Honors Research Project: *A Graphical User Interface for Medical Image Retrieval*
   Supported by the Northeastern University's Matthews Award.
   She is currently pursuing graduate studies at Stanford University.
2. Amanda Funai (Dec. 2004).
   Research Project: *Feature Selection and Clustering Visualization Toolbox*
   Supported by my NSF Career Award.
   She is currently pursuing her Ph.D. degree at University of Michigan, Ann Arbor, and was awarded with an NSF Graduate Research Fellowship in 2007.
3. Michael Mazzello
   Research Project: *Radioisotope Detection*, Supported by CENSSIS.

Ph.D. Thesis Committees:

1. Juan Carlos Rojas (Thesis, Aug. 2003).
   "Multimedia Macros for Portable Optimized Programs." Advisor: Miriam Leeser
2. Anupama Jagannathan (Proposal, Nov. 2004; Thesis, Aug. 2005).
   "Segmentation and Recognition of 3D Point Clouds within Graph-theoretic and Thermodynamic Frameworks." Advisor: Eric Miller
3. Yijian Wang (Proposal, June 2005; Thesis, Dec. 2006). *"Modeling and Acceleration of File-I/O Dominated Parallel Workloads."* Advisor: David Kaeli
4. Ashley Tarokh (Proposal and Thesis, Aug. 2005). *"Shape-Based Methods for Linear Inverse Scattering Problems."* Advisor: Eric Miller
5. Alireza Aliamiri (Proposal and Thesis, Dec. 2006. *"Statistical Methods for Unexploded Ordnance Discrimination."* Advisor: Eric Miller
6. Haiqian Yu (Proposal, Mar. 2005; Thesis, Dec. 2006).
   "Optimizing Data Intensive Window-based Image Processing on Reconfigurable Hardware Boards." Advisor: Miriam Leeser
7. Bing Zhang (Proposal, Aug. 2006; Thesis, April 2007).
   "Discriminative Feature Optimization for Speech Recognition."
   Advisors: John Makhoul and Gene Cooperman
8. Mei Li (Proposal, Dec. 2006; Thesis, Aug. 2007). *"Symbolic Reasoning About Dynamic Systems in Conflict Alert Situations."* Advisor: Mitch Kokar
9. John Oldham (Proposal, Feb. 2006). *"Proteomic Investigation of the Relationship between Primary and Secondary Metabolism in Plants and Plant Cell Culture."*
   Advisor: Carolyn Lee-Parsons

M.S. Thesis Committees:

1. Michael Estlick (Aug. 2002). *"An FPGA Implementation of the K-Means Algorithms for Image Processing."* Advisor: Miriam Leeser
2. Haiqian Yu (Aug. 2003). *"Memory Architecture for Data Intensive Image Processing Algorithms in Reconfigurable Hardware."* Advisor: Miriam Leeser
3. Shawn Miller (April 2004). *"Enabling a Real-time Solution to Retinal Vascular Tracing Using FPGAs."* Advisor: Miriam Leeser
4. Evangelia Komisopoulou (July 2004). *"Clustering Methods For Accurate Background/Foreground Estimation in cDNA Microarray Images."*
   Advisor: Elias Manolakos
5. Anita Thomas (Aug. 2004). *"Value Prediction with Perceptrons."* Advisor: David Kaeli
6. Govindarajan Thattai (Aug. 2004). *"Discriminative Initialization methods to HMM parameter estimation for Speech recognition."* Advisor: John Makhoul
7. Sushanth Dabbiru (Nov. 2004). *"Statistical Modeling for Story Segmentation of Audio Broadcasts."* Advisor: John Makhoul
8. Hardik Virani (April 2005). *"Self-Organizing Feature Maps combined with Ecological Ordination Techniques for Effective Watershed Management."*
   Advisor: Elias Manolakos
9. Guruprasad Saikumar (Aug. 2005). *"MMI Training for Automatic Segmentation of Conversational Telephone Speech."* Advisor: John Makhoul