

1. Active Learning (<http://www.cs.ualberta.ca/~greiner/R/BudgetedLearning>)

A typical learning process starts with a training sample, which is used by a learning process to produce a predictor. The predictor performance, of course, depends critically on that training sample. In *active* learning, however, the learner has the option of extending the training sample, albeit at a cost; these systems must decide, sequentially what information (features or label) of which instance, should be purchased?

My research contributed to this subfield, by providing novel algorithms [C2]¹ (w/PhD: Y Guo); dealing with “labels” that correspond to image segmentations [C17] (w/PhD: A Farhangfar); and going beyond discriminative models, to produce the best generative model [C20] (w/MSc: L Li). That third system is in “budgeted learning” framework, where the learner is given a fixed total budget for its purchases. As the founder of this subfield, I was honored to give the Keynote address in a recent international meeting in this area.

2. Biological Projects: Proteins and Metabolites (<http://www.cs.ualberta.ca/~bioinfo/PA>, <http://www.hmdb.ca/>) Proteome Analyst (PA) is a publicly available, high-throughput, web-based system for predicting various properties of each protein in an entire proteome. PA’s “subcell location” predictor is currently one of the most accurate and most comprehensive such systems. Our team (w/MSc: A Fyshe and Y Liu) recently further improved this system by using the textual information in abstracts describing that protein (and its homologues) [J6]. MSc student B Bostan guided the work to identify which *signaling* pathways a protein belongs to [J10]. In all cases, this involved first theoretical explorations, following by extensive empirical testing on real-world data; publication in top journals demonstrate the world-class performance of these tools.

The team has also been pioneering one of the newest omics (arguably, the one most relevant to diagnosis and treatment): *metabolomics*, which is the systematic study of the small molecules (“metabolites”) within an organism. This involved organizing some of the earliest symposia in this area [C4], then laying out the infrastructure for their computation study, to produce “the most complete and comprehensive curated collection of human metabolite and human metabolism data in the world, [...containing ...] records for more than 2180 endogenous metabolites” [J4]. We then identified, and quantified, the set of metabolites that appear in cerebrospinal fluid [J9] and human serum [J18]. We also connected a person’s metabolic profile with various states, including gender, diurnal variation and age [J3], as well as diseases, including cachexia [J15].

Many of these analyses are based on interpreting NMR spectra, which is very challenging as the peaks associated with each metabolite can shift based on unobserved properties of the mixture. My MSc student M Ravanbakhsh and PDF B Poczos solved this, using a very clever extension to cross-entropy that exploits the observation that the shifts are constrained to a relatively narrow region of the spectrum, which imposes weak constraints on the possible interpretations [C21].

3. Medical-Informatics Projects The following projects are all explicitly related to disease or other medical conditions; all are in collaboration with research and/or clinical physicians. With trainee N Asgarian, we used a machine learning approach to learn a classifier that could accurately predict relapse, based on the *subcellular localization* of certain junctional proteins [J14]. We analyzed the single nucleotide polymorphisms (SNPs) associated with toxicity in patients treated with conformal radiotherapy [J5], and with breast cancer susceptibility [J20]. We (including PDF C Yu) also produced a novel system for predicting survival times for patients – basically a patient-specific Kaplan-Meier curve, using all available patient data, which we show works significantly better than existing techniques [C24].

Several other projects involve medical imaging: [J11] learns to segment tumors within PET scans of a patient’s lungs. We (w/M Morris (MSc), M Schmidt (MSc) and I Diaz (PhD)) have several results that deal with brain tumors: Given a Magnetic Resonance image of a patient’s brain, (1) [C23] describes which pre-processing technique leads to the best (tumor vs healthy) segmentation; and (2) [J19] provides a fast algorithm for approximating the location of the tumor.

¹Each [Jx] and [Cx] reference refers to an entry in the Publication List, which also shows my collaborators.

Awards, Invited Lectures

- **Fellow of AAAI**: (Association for the Advancement of Artificial Intelligence); awarded 2007
- **Faculty Research Award**, UofAlberta, Computing Science, 2007
- **Killam Fellowship**, UofAlberta, 2006–2007.
- **ASTech Award** (to AICML) for “Outstanding Leadership in Technology”, Oct 2006
- **McCalla Professorship**: UofAlberta; 2005–2006
- **Best Paper**: O Schulte, G Frigo, H Khosravi, and R Greiner. (*Canadian Conference on Artificial Intelligence (CAI)*, 2010).
- **Success at Competitions**
 - * Placed 3rd in “Multi-omics challenge on Obstructive Nephropathy” challenge (2010) <http://tunedit.org/challenge/ON>
 - * Obtained the top score in “ADHD-200 Global Competition” (2011) <http://aicml.cs.ualberta.ca/?q=node/80>
- Delivered over 40 invited talks at research labs and universities, in addition to conferences presentations.

Service

- **Scientific Director**, Alberta Ingenuity Centre for Machine Learning, 2006-2007
- **Journal editorial boards**: *Machine Learning*, Springer; *Journal of Artificial Intelligence Research*, AI Access (Associate Editor); *Journal of Machine Learning Research*, AI Access.
- (*Senior*) *Program Committee* for over 20 conferences, including IJCAI, AAAI, UAI, ICML, NIPS
- Member, Evaluation Committee [Discovery Grants], NSERC 1507 (Computer Science); 2010–2013

- [P1] “System and Method for Solving NonLinear Optimization Problems using Cross Entropy Exploiting Partial Decomposability” with Siamak (Mohsen) Ravanbakhsh, Barnabas Poczozos Provisional Patent filed 9 July 2010
- [P2] “Automatic identification of compounds in a sample mixture by means of NMR spectroscopy” with D. Wishart, T. Rosborough, B. Lefebvre, N. Epstein, J. Newton, W. Wong; (7181348; Awarded 20 Feb 2007).
- [P3] “A Method and System for Automatic Detection and Segmentation of Brain Tumors and Associated Edema (Swelling) in Magnetic Resonance Images (MRI)” with M. Schmidt and A. Murtha;
(US Provisional Patent Application filed: 29 April 2005 — 60/675,085)
- [P4] “An Efficient Data-Driven Theory Revision System”
with R.B. Rao and G. Meredith;
(5787232, Awarded 28 July 1998.)
- [P5] “Delta learning system for using expert advice to revise diagnostic expert system fault hierarchies” with R.B. Rao and G. Drastal;
(5987445, Awarded 16 November 1999).
- [P6] “Process, apparatus, media and signals for automatically identifying compounds in a sample”
with D. Wishart, B. Lefebvre, J. Newton, N. Epstein, T. Rosborough, W. Wong
UK Patent: GB2410559
US application filed November 2001.

Refereed journal papers (2007–2012)²

- [J1] **T Van Allen, A Singh**, R Greiner, and P Hooper. Quantifying the uncertainty of a belief net response: Bayesian error-bars for belief net inference. *Artificial Intelligence*, 2007.
- [J2] **L Li, V Bulitko**, and R Greiner. Focus of attention in reinforcement learning. *J Universal Computer Science*, 13(9), 2007.
- [J3] C Slupsky, K Rankin, H Fu, D Chang, E Saude, B Lix, D Adamko, S Shah, R Greiner, B Sykes, and T Marrie. Investigations of the effects of gender, diurnal variation and age in human urinary metabolomic profiles. *Analytical Chemistry*, 2007.
- [J4] D Wishart, D Tzur, C Knox, **R Eisner**, A Guo, N Young, D Cheng, **K Jewell**, D Arndt, S Sawhney, C Fung, L Nikolai, M Lewis, M Coutouly, I Forsythe, P Tang, S Shrivastava, K Jeroncic, P Stothard, G Amegbey, D Block, D Hau, **J Wagner**, J Miniaci, M Clements, M Gebremedhin, N Guo, Y Zhang, G Duggan, G MacInnis, A Weljie, R Dowlatabadi, F Bamforth, D Clive, R Greiner, L Li, T Marrie, B Sykes, H Vogel, and L Querengesser. HMDB: the human metabolome database. *Nucleic Acids Research*, 35, 2007.
- [J5] S Damaraju, B Sehwat, D Carandang, Rashmi Penugonde, R Greiner, and M Parliament. Candidate and whole-genome snp association studies of late radiation toxicity in prostate cancer patients. *Radiation Research*, 170:671–672, 2008.
- [J6] A Fyshe, Y Liu, D Szafron, R Greiner, and P Lu. Improving subcellular localization prediction using text classification and the gene ontology. *Bioinformatics*, 2008.
- [J7] **C Lee**, O Zaiane, H Park, J Huang, and R Greiner. Clustering high dimensional data: A graph-based relaxed optimization approach. *Information Sciences*, 2008.
- [J8] **I Levner**, H Zhang, and R Greiner. Heterogeneous stacking for classification driven watershed segmentation. *EURASIP J Advances in Signal Processing*, 2008(485821), 2008.
- [J9] D Wishart, M Lewis, J Morrissey, M Flegel, Y Xiong, K Jeroncic, D Cheng, **R Eisner**, B Gautam, D Tzur, S Sawhney, F Bamforth, R Greiner, and L Li. The human cerebrospinal fluid metabolome. *J Chromatography B*, 2008.
- [J10] **B Bostan**, R Greiner, D Szafron, and P Lu. Predicting homologous signaling pathways using machine learning. *Bioinformatics*, 2009.
- [J11] A Kerhet, C Small, H Quon, T Riauka, L Schrader, R Greiner, D Yee, A McEwan, and W Rao. Application of machine learning methodology for pet-based definition of lung cancer. *Current Oncology*, 17(1), 2009.
- [J12] O Schulte, W Luo, and R Greiner. Mind change optimal learning of bayes net structure from dependency and independency data. *Information and Computation*, 298, 2009.
- [J13] **X Su**, T Khoshgoftaar, and R Greiner. Making an accurate classifier ensemble by voting on classifications from imputed learning sets. *Int'l J Info & Decision Sciences*, 1(3), 2009.
- [J14] **N Asgarian**, X Hu, Z Aktary, K Chapman, L Lam, R Chibbar, J Mackey, R Greiner, and M Pasdar. Learning to predict relapse in invasive ductal carcinomas based on the subcellular localization of junctional proteins. *Breast Cancer Research and Treatment*, 121(2):527, 2010.
- [J15] **R Eisner**, J Xia, D Hau, **T Eastman**, C Stretch, S Damaraju, R Greiner, D Wishart, and V Baracos. Learning to predict cancer-associated skeletal muscle wasting from ¹H-NMR profiles of urinary metabolites. *Metabolomics*, 7(1):25–34, 2010.
- [J16] **D Lizotte**, R Greiner, and D Schuurmans. An experimental methodology for response surface optimization methods. *J Global Optimization*, 2011.

²The names of students, postdocs and employees under my (co)supervision appear are **boldfaced**. My contributions to every one of these articles was funded, in part, by my previous NSERC grant. This listing does not include 8 other refereed papers, nor 20 posters and invited (but not refereed) publications; see <http://tiny.cc/kby1v>.

- [J17] D Moulavi, **M Hajiloo**, J Sander, P Halloran, and R Greiner. Combining gene expression and interaction network data to improve kidney lesion score prediction. *Int'l J Bioinformatics Research and Applications*, 2011.
- [J18] N Psychogios, D Hau, J Peng, A Guo, R Mandal, S Bouatra, I Sinelnikov, R Krishnamurthy, **R Eisner**, B Gautam, N Young, J Xia, C Knox, E Dong, P Huang, Z Hollander, T Pedersen, S Smith, F Bamforth, R Greiner, B McManus, J Newman, T Goodfriend, and D Wishart. The human serum metabolome. *PLoS One*, 6(2), 2011.
- [J19] B Saha, N Ray, R Greiner, A Murtha, and H Zhang. Quick detection of brain tumors and edemas: A bounding box method using symmetry. *Computerized Medical Imaging and Graphics*, 2011.
- [J20] B Sehrawat, M Sridharan, S Ghosh, P Robson, C Cass, J Mackey, R Greiner, and S Damaraju. Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Human Genetics*, 2011.
- [J21] A Zarnani, P Musilek, X Shi, X Ke, H He, and R Greiner. Learning to predict ice accretion on electric power lines. *Eng. Appl. AI*, 2011.
- [J22] C Stretch, **T Eastman**, R Mandal, **R Eisner**, D Wishart, M Mourtzakis, C Prado, Sambasivarao Damaraju, Ron Ball, R Greiner, and V Baracos. Prediction of skeletal muscle and fat mass in patients with advanced cancer using a metabolomic approach. *J Nutrition*, 2012.
- [J23] **S Wang**, S Wang, L Cheng, R Greiner, and D Schuurmans. Exploiting syntactic, semantic and lexical regularities in language modeling via directed markov random fields. *Computational Intelligence*, 2012.

Refereed Conference Articles (Full paper refereed, under 1-in-3 acceptance rate)³

- [C1] A Farhangfar, R Greiner, and M Zinkevich. A fast way to produce optimal fixed-depth decision trees. In *Int'l Symposium on Artificial Intelligence and Mathematics (AI&Math)*, 2007.
- [C2] **Y Guo** and R Greiner. Optimistic active learning using mutual information. In *Int'l Joint Conference on Artificial Intelligence (IJCAI)(*)*, 2007.
- [C3] O Schulte, W Luo, and R Greiner. Mind change optimal learning of bayes net structure. In *Conference on Learning Theory (COLT)(*)*, 2007.
- [C4] D Wishart and R Greiner. Computational approaches to metabolomics: An introduction. In *Pacific Symposium on Biocomputing*, 2007.
- [C5] **X Su**, R Greiner, T Khoshgoftaar, and X Zhu. Hybrid collaborative filtering algorithms using a mixture of experts. In *IEEE/WIC/ACM Int'l Conf. Web Intelligence (WebIntelligence)*, 2007.
- [C6] **A Isaza**, **J Lu**, Vadim Bulitko, and R Greiner. A cover-based approach to multi-agent moving target pursuit. In *Artificial Intelligence and Interactive Entertainment Conference (AIIDE)(*)*, 2008.
- [C7] **A Isaza**, C Szepesvari, R Greiner, and V Bulitko. Speeding up planning in Markov decision processes via automatically constructed abstractions. In *Uncertainty in Artificial Intelligence (UAI)(*)*, 2008.
- [C8] **C Lee**, **M Brown**, **S Wang**, A Murtha, and R Greiner. Constrained classification on structured data. In *National Conference on Artificial Intelligence (AAAI)(*)*, 2008.
- [C9] **C Lee**, **S Wang**, **M Brown**, A Murtha, and R Greiner. Segmenting brain tumors using pseudo-conditional random fields. In *Medical Image Computing and Computer-Assisted Intervention (MICCA)(*)*, 2008.
- [C10] **I Levner**, R Greiner, and H Zhang. Supervised image segmentation via ground truth decomposition. In *IEEE Int'l Conf. Image Processing*, 2008.

³All of the following Computing Science conferences are archival venues, and serve as the primary means for disseminating results; the ones marked with a (*) are especially prestigious.

- [C11] J Lees-Miller, F Anderson, **B Hoehn**, and R Greiner. Does Wikipedia information help Netflix predictions? In *Int'l Conference on Machine Learning and Applications (ICMLA)*, 2008.
- [C12] **X Su**, T Khoshgoftaar, and R Greiner. Imputed neighborhood based collaborative filtering. In *IEEE/WIC/ACM Int'l Conf. Web Intelligence (WebIntelligence)*, 2008.
- [C13] **X Su**, T Khoshgoftaar, and R Greiner. A mixture imputation-boosted collaborative filter. In *Florida AI Research Symposium (FLAIRS)*, 2008.
- [C14] **X Su**, T Khoshgoftaar, and R Greiner. Using imputation techniques to help learn accurate classifiers. In *Fifteenth IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*, 2008.
- [C15] **X Su**, T Khoshgoftaar, X Zhu, and R Greiner. Imputation-boosted collaborative filtering using machine learning classifiers. In *ACM Symposium on Applied Computing (ACM Applied Comput.)*, 2008.
- [C16] P Hooper, Y Abbasi-Yadkori, R Greiner, and **B Hoehn**. Improved mean and variance approximations for belief net responses via network doubling. In *Conference on Uncertainty in Artificial Intelligence (UAI)(*)*, 2009.
- [C17] **A Farhangfar**, R Greiner, and C Szepesvari. Learning to segment from a few well-selected training images. In *International Conference on Machine Learning (ICML)(*)*, 2009.
- [C18] **B Póczos**, Y Abbasi-Yadkori, C Szepesvari, R Greiner, and N Sturtevant. Learning when to stop thinking and do something! In *Int'l Conference on Machine Learning (ICML)(*)*, 2009.
- [C19] O Schulte, G Frigo, R Greiner, and H Khosravi. A new hybrid method for bayesian network learning. In *Symp. Computational Intelligence and Data Mining*, 2009.
- [C20] **L Li**, **B Póczos**, C Szepesvari, and R Greiner. Budgeted distribution learning of belief net parameters. In *Int'l Conference on Machine Learning (ICML)(*)*, 2010.
- [C21] **S Ravanbakhsh**, **B Póczos**, and R Greiner. A cross-entropy method that optimizes partially decomposable problems: A new way to interpret NMR spectra. In *National Conference on Artificial Intelligence (AAAI)(*)*, 2010.
- [C22] O Schulte, G Frigo, H Khosravi, and R Greiner. The IMAP hybrid method for learning gaussian bayes nets. In *Canadian Conference on Artificial Intelligence (CAI)*, 2010. Awarded Best Paper prize
- [C23] **I Diaz**, P Boulanger, R Greiner, and A Murtha. A critical review of the effect of de-noising algorithms on MRI brain. In *IEEE Medicine and Biology Society*, 2011.
- [C24] **C Yu**, R Greiner, **H Lin**, and V Baracos. Learning patient-specific cancer survival distributions as a sequence of dependent regressors. In *Neural Information Processing Systems (NIPS)(*)*, 2011.
- [C25] **X Su**, T Khoshgoftaar, and R Greiner. Vipboost: a more accurate boosting algorithm. In *Florida AI Research Symposium (FLAIRS)*, 2009.
- [C26] **X Su**, R Greiner, T Khoshgoftaar, and A Napolitano. Using classifier-based nominal imputation to improve machine learning. In *Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2011.