Editorial

Leif E. Peterson

Center for Biostatistics,
The Methodist Hospital Research Institute,
6565 Fannin, Suite MGJ6-031,
Houston, Texas 77030, USA
E-mail: lepeterson@tmhs.org

Xue-Wen Chen

Information and Telecommunication Technology Center,
Bioinformatics and Computational Life Sciences Laboratory,
Electrical Engineering and Computer Science Department,
The University of Kansas,
1520 W. 15th Street, 2028 Eaton,
Lawrence, Kansas 66045, USA
E-mail: xwchen@ku.edu

Biographical notes: Leif E. Peterson received the BSc Degree in Nuclear Medicine from Ferris State University in 1981, MPH Degree in Radiological Physics from University of Michigan in 1986, and PhD Degree in Epidemiology and Biostatistics from University of Texas in 1994. He is currently Associate Professor of Public Health (pending), Weill Medical College, Cornell University, and Director of the Center for Biostatistics, The Methodist Hospital Research Institute. His research interests include machine learning and computational intelligence in bioinformatics, oncology, and cardiovascular disease.

This special issue is comprised of papers that are extended version of papers presented at *ICMLA 2007* on 'Machine Learning in Biomedicine and Bioinformatics'. The focus of the session was to introduce new machine learning problems and attract newly developed machine learning methods in biomedicine and bioinformatics. In addition, the goal of the session was to foster new research collaborations between the biomedicine/ bioinformatics community and the machine learning community.

The dramatic growth of medical and biological data has created an unprecedented opportunity for machine learning in the pattern recognition and machine learning community. Many medical and biological problems involve challenging approaches to pattern discovery and learning. Increasingly, intelligent systems have been designed to solve problems in biology and medicine. In the main, the major computational challenges in machine learning applications for biomedicine and bioinformatics include the small sample problem ($n \ll p$) or curse of dimensionality, missing data and imputation-validation, finding cluster structures of data, motif finding, proton-protein

interaction network development, discerning complex decision boundaries, integration and information retrieval from disparate data sets.

Papers included in this special issue cover sequence analysis, gene identification, classifier ensemble methods, protein-protein interaction network, protein searching with a database-derived distance constraints, and protein residue torsion angle distributions. The paper by Chen et al. addresses DNA sequence motif analysis using knowledge-driven multilevel ICA. Classifier performance and ensemble methods using random spherical and principal direction linear oracles are addressed in the paper by Peterson and Coleman. Gene identification and survival prediction with Lp Cox regression and novel similarity measure are addresses in the paper by Liu and Jiang. The paper by Taylor et al. covers an analysis pipeline for the inference of protein-protein interaction networks. Another paper by Fionda et al., covers a novel search techniques for functional similarities in protein-protein interaction networks. Refinement of under-determined loops of human prion protein by database-derived distance constraints are covered in the paper by Cui et al. Lastly, the paper by Sun et al. introduces PRTAD: a database for protein residue torsion angle distributions.