VASANT G. HONAVAR

Dr. Vasant Honavar
Professor and Edward Endowed Professor
vhonavar@psu.edu
College of Information Science and Technology
8148653141
E335 Westgate Building
Pennsylvania State University
http://faculty.ist.psu.edu/vhonavar

CITIZENSHIP

US Citizen through Naturalization

EDUCATION

University of Wisconsin-Madison, USA	Ph.D. Computer Science and Cognitive Science (Advisor: Leonard Uhr)	1990
University of Wisconsin-Madison, USA	M.S. Computer Science	1989
Drexel University	M.S. Electrical and Computer Engineering	1984
Bangalore University	B.E. Electronics Engineering	1982

PRIMARY TEACHING AND RESEARCH INTERESTS

Artificial Intelligence, Machine Learning, Causal Inference, Bioinformatics and Computational Biology, Discovery Informatics, Big Data Analytics, Health Informatics, Neuroinformatics, Materials Informatics, Social Informatics, Knowledge Representation and Inference, Semantic Web.

ACADEMIC APPOINTMENTS

2013-	Professor of Informatics	Penn State Univ.
2013-	Edward Frymoyer Endowed Endowed Professor of IST	Penn State Univ.
2015-	Professor of Computer Science and Engineering	Penn State Univ.
2016-	Professor of Data Sciences	Penn State Univ.
2013-	Director, Artificial Intelligence Research Laboratory	Penn State Univ.
2015-	Associate Director, Institute for Computational and Data Sciences	Penn State Univ.
2015-	Director, Center for Big Data Analytics and Discovery Informatics	Penn State Univ.
2016-	Sudha Murty Distinguished Visiting Chair in Neurocomputing & Data Science, School of Automation	Indian Institute of Science
2001-13	Professor of Computer Science	Iowa State Univ.
2002	Visiting Professor, Medical Informatics & Biological Statistics	Univ. of Wisconsin
1996-01	Associate Professor, Computer Science	Iowa State Univ.
1998	Visiting Professor, Computer Science	CMU
1990-96	Assistant Professor, Computer Science	Iowa State Univ.
1986-90	Research Assistant, Computer Science	Univ. Wisconsin
1984-86	Teaching Assistant, Electrical and Computer Engineering	Univ. Wisconsin
1982-84	Teaching Assistant, Electrical and Computer Engineering	Drexel Univ.

LEADERSHIP EXPERIENCE

2020-	Cross-Hub Coordination Committee	NSF Big Data Hubs
2020-	AI and Blockchain Task Force	MDEpinet

2016-	Co-director, Biomedical Data Sciences PhD Training Program	Penn State Univ.
2016-17	Chair, Data Sciences Faculty Search, College of IST	Penn State Univ.
2016-19	Co-Director, Data Sciences Undergraduate Program, College of IST	Penn State Univ.
2014-	Associate Director, Institute for Computational and Data Sciences	Penn State Univ.
2015-	Director, Center for Big Data Analytics and Discovery Informatics	Penn State Univ.
2015-	Co-director, Informatics, Clinical and Translational Sciences Institute	Penn State Univ.
2014-18	Council Member, Computing Community Consortium	CRA
2015-20	Executive Committee Member	NSF NE Big Data Hub
2010-13	Program Director, Information and Intelligent Systems Division	NSF
2005-13	Director, Ctr. for Computational Intelligence, Learning, & Discovery	Iowa State Univ.
2006-10	Director of Research, Department of Computer Science	Iowa State Univ.
2009-10	Chair, Promotion and Tenure Committee, Computer Science	Iowa State Univ.
2003-05	Chair, Bioinformatics & Computational Biology Graduate Program	Iowa State Univ.
2003-04	Chair, Research Space Committee, Computer Science	Iowa State Univ.
2002-10	Director of Research, Computer Science	Iowa State Univ.
2001-02	Chair, Graduate Admissions Committee, Computer Science	Iowa State Univ.
2001-03	Associate Chair, Bioinformatics & Comp. Biol. Graduate Program	Iowa State Univ.
1999-01	Chair, Graduate Admissions, Computer Science	Iowa State Univ.
2001	Co-Chair, Strategic Planning, Computer Science	Iowa State Univ.
1999-00	Chair, Graduate Admissions, Bioinformatics & Comp. Biol. Program	Iowa State Univ.
1999-00	Chair, Faculty Search, Computer Science	Iowa State Univ.
1990-13	Founder and Head, Artificial Intelligence Research Laboratory	Iowa State Univ.

INTERDEPARTMENTAL AND INTERDISCIPLINARY PROGRAM PARTICIPATION

2013-	The Huck Institutes of the Life Sciences	Penn State Univ.
2013-	Institute for Cyberscience	PennState Univ.
2014-	Bioinformatics and Genomics Graduate Program	Penn State Univ.
2014-	Operations Research Graduate Program	Penn State Univ.
2014-	Neuroscience Graduate Program	Penn State Univ.
2004-13	Center for Computational Intelligence, Learning, & Discovery	Iowa State Univ.
1999-13	Bioinformatics & Computational Biology Graduate Program	Iowa State Univ.
1999-	Laurence H. Baker Center for Bioinformatics and Biological Statistics	Iowa State Univ.
1999-13	Computational Molecular Biology Training Group	Iowa State Univ.
2002-05	Comp. Biol. for Animal Agriculture Training Group	Iowa State Univ.
2003-13	Human Computer Interaction Graduate Program	Iowa State Univ.
2001-13	Information Assurance Masters Program	Iowa State Univ.
2001-13	Information Assurance Center	Iowa State Univ.
2002-13	Institute for Science and Society	Iowa State Univ.
1992-13	Neuroscience	Iowa State Univ.

My primary responsibilities at Iowa State University during 1990-2010 were Research (50%), Teaching (50%), and Service. During 2010-2013, I have been on an IPA assignment as a Program Director at NSF while maintaining my research activities at ISU. In September 2013, I joined Pennsylvania State University where my primary responsibilities are Research (50%), Teaching (50%) and Service.

CAREER HIGHLIGHTS

- **Pennsylvania State University** (2013-present)
  - **Edward Frymoyer Endowed Professor of Information Sciences and Technology:**
    - **Artificial Intelligence Research Laboratory**: Led the development of Artificial Intelligence Research Laboratory as a focal point for research in Artificial Intelligence (especially Machine Learning, Causal Inference, Big Data Analytics, Knowledge Representation and applications in Bioinformatics, Health Informatics, and Security Informatics.
    - **Research and Graduate and Postdoctoral Training:** Conceiving and leading research projects in Machine Learning (especially predictive modeling of high-dimensional, longitudinal data, network data), Big Data Analytics, Knowledge Representation (especially secrecy-preserving query answering and preference reasoning) and Causal Inference (especially eliciting causal effects from temporal and relational data), and applications in bioinformatics (especially computational prediction of protein-protein and protein nucleic acid interfaces and complexes, analysis of microbiomes), health informatics (e.g., predictive modeling of health outcomes from clinical, biomedical, environmental, and socio- economic data), and data and computing infrastructure for data-intensive interdisciplinary collaborations, funded in part by grants from the National Science Foundation, the National Institutes of Health, and National Security Agency. Graduated 3 PhD students and currently supervising a research group consisting of 8 PhD Students (5 Informatics, 2 in Computer Science and Engineering, 1 in Bioinformatics and Genomics).
    - **Undergraduate and Graduate Programs in Data Sciences:** Co-led the establishment of a new innovative inter-college undergraduate degree program in Data Sciences, that combines foundational training in Computing, Informatics, and Statistics (e.g., Statistical Methods, Databases and Data Integration, Programming and Data Structures, Machine Learning, Big Data Algorithms, Programming Models for Big Data, Capstone Project) together with exposure to least one application domain (e.g., Life Sciences, Social Sciences, Business, Health Sciences, Physical Sciences). Developed or co-developed Data Sciences undergraduate courses in Machine Learning, and related topics); Led the design and development of an innovative Masters program in Data Sciences (currently in the process of being formalized), co-led the development of an AI undergraduate program.
    - **Service to the College, University**: Institute for Computational and Data Science (ICDS) Associate Director 2014-present, ICDS Faculty Council Member 2020-present, Member of the Data Sciences coordination committee (2020-present), Chair of the Data Sciences Faculty Search Committee (2016-2017), Graduate Recruitment Committee (2016-2017), Faculty Council (2016-2021), Graduate Advisory Committee (2014-2016, 2018-2020), Promotion and Tenure Committee (2016-17), Faculty Evaluation Committee (Research) (2016-2021), Strategic Planning Committee (2015- 2016). Led the Data Sciences thrust of the IST strategic plan (2016); Led a major faculty hiring initiative in the Data Sciences (including co-hires in Big Data Life Sciences in coordination with the Huck Institutes of the Life Sciences); Contributed to the "Driving Digital Innovation" and "Precision Health in Context" thrusts of the University's strategic plan. Mentored junior faculty within and outside the college, e.g., in developing successful NSF CAREER proposals, NIH K proposals, etc.

Service to the Discipline and the Scientific Community: As a member of **Computing Community Consortium Council** (2014-2017), a standing committee of the Computing Research Association, charged by NSF with developing and articulating computing research agendas organized

around emerging opportunities or national or societal challenges and priorities, chaired the CCC task force on Convergence of Data and Computing, contributed to task forces on Artificial Intelligence and on Health IT, organized the AAAI/CCC Symposium on "Accelerating Science: A Grand Challenge for AI", led the CCC white paper on "Accelerating Science: Computing Research Agenda", contributed to several CCC white papers and research agenda setting workshops, co-organized the CCC Symposium on Computing Research: Addressing National Priorities and Societal Needs. As a member of the Electorate Nominating Committee (ENC) of the Information, Computing, and Communication Section (Section T) of the American Association for Advancement of Science (AAAS), helped identify candidates for key leadership positions within AAAS Section T.

- **Founding Director, PSU Center for Big Data Analytics and Discovery Informatics (2015-present):**
  - **Center Leadership:** Led the development of the Center for Big Data Analytics and Discovery Informatics co-sponsored by the College of Information Sciences and Technology, Institute for Computational and Data Sciences, the Huck Institutes of the Life Sciences, and the Social Science Research Institute that brings together faculty with expertise in computer science, informatics, and statistics with researchers across several scientific domains (Health, Life Sciences, Brain Sciences, Environmental Sciences, Social Sciences, Material Sciences) to pursue foundational research on computational and statistical methods, tools, and data and computing infrastructure and their application to problems of societal importance. Leading an NSF-funded planning an NSF AI Institute for Materials Discovery; Leading a university Center for Scientific Artificial Intelligence that brings together expertise in Artificial Intelligence and Scientific applications across the university to pursue foundational and applied research in AI focused on accelerating science.
  - **Penn State Clinical and Translational Sciences Institute (CTSI)**: Informatics co-lead of Penn State CTSI (an NIH-funded 4-year project with a budget close to $20 million, PI: Lawrence Sinoway) focused on developing data and computational infrastructure including data access and use policy compliant integration and analyses of clinical, biomedical, socio-demographic, environmental, behavioral, and other types of data for predictive and causal modeling of health outcomes and personalization of interventions.
  - **Virtual Data Collaboratory:** Leading (with Manish Parashar of Rutgers University) the design, implementation, and evaluation of a multi-university distributed cyberinfrastructure for collaborative, data-intensive science; and co-leading (with Helen Berman of Rutgers University) a life sciences use case focused on novel methods and distributed workflows for analyses and prediction of protein nucleic acid interactions, interfaces, and complexes (An NSF funded $4 million, 4 year project).
  - **Penn State Biomedical Data Sciences (BD2K) Doctoral Training Program:** Co-led the establishment of a new NIH-funded innovative PhD training program in Biomedical Data Sciences to prepare a diverse cadre of Biomedical Data Scientists with deep expertise in computing, statistics, and informatics and adequate exposure to some area of Biomedical and Health Sciences develop and apply innovative algorithms and statistical methods for predictive and causal modeling of, and effective interventions that impact individual and population health outcomes through integrative analysis of clinical, biomedical, behavioral, socio-economic, environmental, and other types of data.
  - **Interdisciplinary Study Groups**: Initiated interdisciplinary study groups to foster collaborative research at the intersection of Data Sciences and (i) Health Sciences; (ii) Life Sciences; (iii) Brain Sciences.

- **Northeast Big Data Innovation Hub:** As member of the Executive Committee and Co-PI of the Northeast Big Data Innovation Hub, an NSF-funded consortium of academic, industrial, non- profit

organizations aimed at fostering translational data sciences research, education, and workforce development aimed at addressing complex national and regional challenges (e.g., accelerating science, improving health, enabling smart cities and communities), worked with other members of the Hub leadership to develop, articulate, and implement the Hub's vision, strategic priorities, activities. Co-led a successful NSF funded proposal for a 'spoke' on integration and sharing of population health data for data-intensive health research and novel data analytics methods development and evaluation.

- **Eastern Regional Network:** As a member of the steering committee of the Eastern Regional Network, a consortium of universities in the Eastern region of the United States organized around developing innovative computational and data infrastructure for lowering the barriers to transformative collaborative research that transcends institutional and disciplinary boundaries, conceived and co-led the planning activities, with support from National Science Foundation.
  - **Associate Director, PSU Institute for Computational and Data Science (ICDS)**
    - Contributed to developing the vision, mission, strategic plan, and priorities of Institute for Computational and Data Sciences, a university-wide institute aimed at fostering research in data and computation intensive discovery across all areas of science and engineering.
    - Served ICDS in a number of capacities including representing ICDS in strategic co-hiring of faculty with a number of academic units across the university (e.g., in Data Sciences, Biomedical Sciences, and related areas); representing ICDS in campus-wide initiatives e.g., in Precision Health, Microbiome, Data Sciences; Representing ICDS in strategic planning related activities; Organizing seed grant competitions; Mentoring junior faculty; Fostering interdisciplinary collaborative research; Developing advanced research computing infrastructure, etc.
    - Led the development of the Scientific Artificial Intelligence initiative, a faculty-led research initiative of the ICDS aimed at elevating Penn State to a position of national leadership in developing and applying cutting edge Artificial Intelligence and machine learning methodologies to advance a number of scientific disciplines, including materials discovery, predictive and causal modeling of health risks and health outcomes, understanding molecular structure, function, and dynamics, among others.
- **National Science Foundation** (2010-2013).
  - **BIGDATA: Critical Techniques and Technologies for Advancing Big Data Science and Engineering (2011-2013).** Lead Program Director. Responsible for drafting/revising solicitation, managing program budget, coordination with multiple NSF directorates and NIH institutes, organizing and managing multiple review panels, making funding recommendations, managing awards, communicating with NSF leadership and the broader scientific community, Participating in Interagency Big Data R&D Working Group coordinated by NITRD.
  - **III: Information Integration and Informatics Core Program (2010-2013)**. Program Director, Responsible for contributing to the solicitation, organizing and managing multiple review panels for small, medium, large and CAREER proposals, making funding recommendations, managing awards, on a variety of topics related to Data Mining, Machine Learning, Social Networks and Social Media, Information Integration, Bioinformatics, and Semantic Technologies
  - **Expeditions in Computing (2010-2013).** Program Director, co-managing review panels, providing input on funding recommendations, managing 2 expeditions awards (Making Sense at Scale with Algorithms, Machines and People, Understanding Climate Change: A Data Driven Approach), organizing annual reviews and site visits, communicating with NSF leadership and the broader scientific community.
  - **SHB: Smart Health and Wellbeing. (2011-2013).** Program Director, Responsible for contributing to the solicitation, organizing and managing multiple review panels, making funding recommendations, managing awards

- **SLC: Science of Learning Centers (2011-2013).** CISE representative on the NSF oversight team for the Pittsburgh Science of Learning Center, responsible for project oversight, annual review and recommendations.
- **IGERT: Integrative Graduate Education and Research Training (2011-2013).** Program Directors, Responsible for organizing proposal review panels and making funding recommendations on CISE-IIS-relevant proposals.
- **CRI: Computing Research Infrastructure (2011-2013):** Program Director, Responsible for organizing proposal review panels and making funding recommendations for CISE-IIS-relevant proposals.
- **NRI: National Robotics Initiative (2011-2013).** Program Director, Responsible for organizing proposal review panels and making funding recommendations for CISE-IIS-relevant proposals.
- **SI^2: Software for Sustained Innovation**, Program Director, Responsible for co-organizing panels and making funding recommendations for CISE-relevant proposals.
- **Research Planning Workshops:** Initiated research planning workshops on
  - Discovery Informatics (organized by Yolanda Gil and Haym Hirsh),
  - Population Health Measurement and Analysis (organized by Bruce Schatz)
  - Finance Informatics (organized by Louiqa Raschid)
  - Knowledge Representation (organized by Natasha Noy and Deborah McGuinness).
- **NSF-OFR inter-agency collaborative research initiative in Finance Informatics:** Program Director. Responsible for a Dear Colleague Letter and Memorandum of Understanding between NSF and OFR on a joint NSF-OFR initiative focused on algorithms, informatics, knowledge representation, and data analytics needed to advance the current state of the art in financial research and analysis.

- **Iowa State University**
  - **Professor of Computer Science and of Bioinformatics and Computational Biology** (1990-2013)
    - **Research:** Foundational contributions in Machine learning, Data Mining, Knowledge Representation and Semantic Web and interdisciplinary contributions in Bioinformatics and Computational Biology: Scalable approaches to learning predictive models from very large, distributed, data; Methods for learning predictive models from richly structured data (including tabular, sequence, network, relational, linked open, multi-modal, time series data); Learning regular and stochastic context free grammars; Eliciting causal information from multiple sources of observational and experimental data; Selective sharing of knowledge across disparate knowledge bases, including novel approaches to secrecy- preserving query answering; Representing and reasoning about preferences; Composing complex software services from components; and applications in bioinformatics and computational molecular and systems biology, including methods and software for characterization, analysis, and prediction of sequence and structural correlates of protein- protein, and protein-RNA interfaces and interactions, comparative analyses of biological networks (http://ailab.ist.psu.edu/software.html).
    - **Teaching and Curriculum Development:** Established, and over the years, refined and broadened graduate and undergraduate curricula, integration of research into undergraduate education in Artificial Intelligence and Machine Learning.
    - **Artificial Intelligence Laboratory:** Established and led Artificial Intelligence Research Laboratory as a focal point for AI research at Iowa State University. The laboratory secured over $20 million in research funding from NSF, NIH, USDA, and DOD trained 30 PhD students

(10 employed as tenure-track or tenured faculty in academia, 10 employed in academic research, and 10 employed in industrial research and development), 25 MS graduates (24 employed in industry, one in academic research).

- **Center for Computational Intelligence, Learning and Discovery (2005-2013)**. Established and led an interdisciplinary research center focused on foundational research in Computer Science (AI, Machine Learning, Informatics, Theoretical Computer Science, Algorithms, Systems), Statistics, and application areas such as Life Sciences, Agriculture, Engineering.
- **Interdepartmental PhD Program in Bioinformatics and Computational Biology:** Co-led the establishment of the interdepartmental graduate program in Bioinformatics and Computational Biology, with support from an NSF Integrative Graduate Education and Research Training (IGERT) award in 2000, renewed for a second 5-year period in 2005, and served as associate chair (2001-2003) and chair (2003-2005) of the of the program during its formative years. Lead a major curriculum development effort focused on required core courses for the graduate program in Bioinformatics and Computational Biology, culminating in the creation of a Bioinformatics graduate core curriculum comprising of 4 core courses covering Genomics to Systems Biology. By 2005, the program had established itself as one of the top graduate programs in Bioinformatics among institutions without a medical school.
- **Undergraduate Program in Bioinformatics and Computational Biology**: Co-led the development of an undergraduate curriculum and the establishment of interdepartmental undergraduate degree program in Bioinformatics and Computational Biology in 2007.
- **Service to the Department, College, and University**: Served the Department of Computer Science in several important roles including Director of Research (2006-2011) responsible for research infrastructure and fostering research collaborations with other disciplines; Committees focused on Strategic Planning, Faculty hiring, Promotion and tenure, Graduate Admissions, Graduate Curriculum, among others; university research computing advisory committee, among others.
- **Service to the Discipline and the Scientific Community:** Service on NIH Study Sections, including Charter member of NIH study section on Biological Data Management and Analysis (2004-2007), NSF Proposal Review panels, Editorial boards of several journals, Organization of several scientific conferences and workshops.

BIOGRAPHICAL SUMMARY

Dr. Vasant Honavar received his Ph.D. in Computer Science and Cognitive Science in 1990 from the University of Wisconsin Madison, specializing in Artificial Intelligence. In September 2013, Honavar joined the faculty of Penn State University where he currently serves as a Professor and Edward Frymoyer Chair of Information Science and Technology. He is also the founding Director of the Center for Big Data Analytics and Discovery Informatics, Associate Director of the Institute for Computational and Data Sciences, Co-Director of the NIH-funded Biomedical Data Sciences PhD program, and Informatics lead (research) for the NIH- funded Clinical and Translational Sciences Institute, Co-PI of the Northeast Big Data Hub, PI of the NSF planning grant for an Institute for AI-enabled Materials Design, Discovery, and Synthesis, and Co-PI of the NSF Eastern Regional Network Midscale Infrastructure planning project. Honavar serves on the faculty of the Computer Science, Bioinformatics and Genomics, Neuroscience, and the Operations Research Graduate Programs at Pennsylvania State University. In 2016, Honavar was appointed as the Sudha Murty Distinguished Visiting Chair of Neurocomputing and Data Science at the Indian Institute of Science. Honavar served as a member of the Computing Community Consortium Council (CCC) where he chaired the Convergence of Data and Computing Task Force (2015-2017) and served on Artificial Intelligence and Health IT task forces, and as a member of the FDA's MDEpinet

Blockchain and AI Taskforce (2020-21). Dr. Honavar served on the Executive Committee of the NSF North East Big Data Innovation Hub (2016-2020) and currently serves on the Cross-Hub Coordination Committee. During 2015-2018 he served as a member of the Electorate Nominating Committee of the Section on Information, Computing, & Communication of the American Association for the Advancement of Science (AAAS). During 2022, Honavar will serve as the program co-chair of the AAAI Conference on Artificial Intelligence. In addition to research, graduate student supervision and teaching, he is responsible for developing new research and educational initiatives in Data Sciences and contributing to research initiatives in Biomedical and Health Sciences.

From 1990 to 2013, Honavar served on the faculty of Computer Science and of Bioinformatics and Computational Biology at Iowa State University (ISU). At ISU, he directed the Artificial Intelligence Research Laboratory (which he founded in 1990) and the Center for Computational Intelligence, Learning & Discovery (which he founded in 2005) and served as the associate chair (2001-2003) and chair (2003-2005) of the ISU Bioinformatics and Computational Biology Graduate Program, which he helped establish in 1999 with support from an Integrative Graduate Education and Research Training (IGERT) award. During 2010-2013, Honavar served as a program director in the Information and Intelligent Systems Division of the Computer and Information Sciences and Engineering directorate of the National Science Foundation (NSF) during 2010-2013 while maintaining his research program at ISU. He led the Big Data Science and Engineering Program, established the NSF-OFR collaboration in Computational and Information Processing Approaches to and Infrastructure in support of, Financial Research and Analysis and Management, contributed to Smart and Connected Health, Information Integration and Informatics, Expeditions in Computing, Science of Learning Centers, Integrative Graduate Education and Research Training, Computing Research Infrastructure Programs.
Honavar's current research and teaching interests include Artificial Intelligence, Machine Learning, Bioinformatics, Causal Inference, Big Data Analytics, Computational Molecular Biology, Discovery Informatics Information Integration, Knowledge Representation and Inference, Biomedical and Health Informatics, Materials Informatics, Social Informatics, Security Informatics. Honavar has led research projects funded by NSF, NIH, and USDA that have resulted in foundational research contributions (documented in over 250 peer-reviewed publications, that have been cited over 15,500 times, or over 500 citations on average per year during 1990-2021) in: scalable approaches to learning predictive models from very large, richly structured data (including tabular, sequence, network, relational, time series data); Eliciting causal information from multiple sources of observational and experimental data, Selective sharing of knowledge across disparate knowledge bases, including novel approaches to secrecy- preserving query answering; Representing and reasoning about preferences; Composing complex software services from components; and applications in bioinformatics and computational molecular and systems biology (including characterization, analysis, and prediction of sequence and structural correlates of protein- protein, and protein-RNA interfaces and interactions, comparative analysis of biological networks (network alignment)). This research has resulted in a broad range of software and tools (See http://ailab.ist.psu.edu/software.html). Honavar's current research is focused on the theoretical foundations and algorithms for eliciting causal effects from relational and temporal data; for machine learning algorithms for predictive modeling from large, heterogeneous, high dimensional, richly structured, longitudinal data, with applications in precisión health, behavioral and cognitive sciences, and related areas: algorithms for modeling and comparative analyses of brain networks structure and dynamics from fMRI, diffusion tensor imaging and other data; development of a distributed computing and data infrastructure for collaborative data intensive research, analyses and prediction of protein- nucleic acids interactions, interfaces and complexes, analysis and modeling of microbiomes. Honavar has served as a principal or co-principal investigator on grants totaling over $60 million during 1990-2020 from the National Science Foundation, the National Institutes of Health, the US Department of Agriculture, the US Department of Defense, and the National Security Agency. He has extensive curriculum development and teaching experience in Artificial Intelligence, Machine Learning, and Bioinformatics. He also has substantial industrial consulting experience in Data Mining, Bioinformatics, and related topics.

Honavar has served on, or currently serves on the editorial boards of several journals including IEEE/ACM

Transactions on Computational Biology and Bioinformatics, Springer Open Journal of Big Data, Journal of Computational Systems Biology, Cognitive Systems Research, Machine Learning, the Journal of Bioinformatics and Biology Insights, the International Journal of Semantic Web and Information Systems, the International Journal of Computational Biology and Drug Design, the International Journal of Computer and Information Security, and the International Journal of Data Mining and Bioinformatics among others. Honavar served as a general co-chair of the 2014 IEEE International Conference on Big Data in 2014 and the program co-chair of 2014 IEEE Conference on Bio and Medical Informatics. Honavar has served on the program committees of major research conferences in artificial intelligence, data mining, and bioinformatics including the Conference on Artificial Intelligence (AAAI), International Conference on Machine Learning (ICML), ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), SIAM Conference on Data Mining (SDM), IEEE Conference on Data Mining (ICDM), Intelligent Systems in Molecular Biology (ISMB), ACM Conference on Bioinformatics and Computational Biology (ACM-BCB), among others. Honavar has served on multiple NIH Study Sections and including a charter member of the National Institutes of Health study section on Biological Data Management and Analysis (2002-2007). Honavar is a senior member of the Association for Computing Machinery (ACM), the Institute of Electrical and Electronic Engineers (IEEE) and the Association for Advancement of Artificial Intelligence (AAAI); and a member of International Society for Computational Biology (ISCB) Society for Industrial and Applied Mathematics, and the American Association for the Advancement of Science. He currently serves on the Board of Directors for ACM Special Interest Group on Bioinformatics.

Honavar has received many awards and honors during his career including election as Fellow of the American Association for Advancement of Science (AAAS), the National Science Foundation Director's Award for Superior Accomplishment in 2013 for his leadership of the NSF Big Data Program, the National Science Foundation Director's Award for Collaborative Integration in 2011, the Pennsylvania State University College of Information Sciences and Technology Senior Faculty Excellence in Research Award in 2016, the Sudha Murty Distinguished Visiting Chair in Neuroscomputing and Data Science at the Indian Institute of Science in 2016, the Edward Frymoyer Endowed Chair in Information Sciences and Technology at Pennsylvania State University in 2013, Iowa Board of Regents Award for Faculty Excellence in 2007, the Iowa State University College of Liberal Arts and Sciences Award for Career Excellence in Research 2008, and the Iowa State University Margaret Ellen White Graduate Faculty Award in 2011. However, his proudest accomplishments are the 34 PhD students, 32 MS Students, and many undergraduate researchers that he has worked with and mentored during his career.

HONORS AND AWARDS

2018	American Association of Advancement of Science (AAAS) Fellow
2018	ACM Distinguished Member Award
2016-	Sudha Murty Distinguished Visiting Chair in Neurocomputing and Data Science, IISc
2016	125 People of Impact, Dept. of Electrical and Computer Engg., Univ. Wisconsin
2016	Senior Faculty Research Excellence Award, College of IST, Penn State Univ.
2014-17	Council Computing Community Consortium, Computing Research Assoc.
2014	Senior Member, Association for the Advancement of Artificial Intelligence
2013-	Edward Frymoyer Endowed Professorship in Information Sciences and Technology
2013	National Science Foundation Director's Award for Superior Accomplishment
2013	Best Student Paper Award (PhD student Harris Lin), IEEE Big Data Congress, 2013
2012	National Science Foundation Director's Award for Collaborative Integration
2011	Margaret Ellen White Graduate Faculty Award, Iowa State University
2011	Best poster award, ACM Conference on Bioinformatics and Computational Biology
2011	Elected Fellow, International Society for Intelligent Biological Medicine
2008	Liberal Arts and Sciences Award for Excellence in Research, Iowa State University
2007	Regents Award for Faculty Excellence, Board of Regents, Iowa

2006	Best Paper Award, Asian Semantic Web Conference
2006	Best Paper Award, IEEE ICTAI 2006
2006	Senior Member, Association for Computing Machinery
2001	Senior Member, IEEE
1994-99	Research Initiation Award, National Science Foundation
1994	Who's who in Science and Engineering
1992	Elected Member, Sigma Xi
1990-	Associate, Behavior and Brain Sciences
1990	Elected Member, New York Academy of Sciences
1990	Fellow, Workshop on Human and Machine Cognition
1989	Fellowship, Summer Institute in Parallel Computing, Argonne National Laboratory
1989	Student Fellowship, International Joint Conference on Artificial Intelligence (IJCAI)
1989	Fellowship, McDonnell Summer Institute in Cognitive Neuroscience, Dartmouth
1988	Fellowship, Connectionist Models Summer School, Carnegie Mellon University
1982	Gold medal for academic excellence, Bangalore University, India
1975-82	National Merit Scholar, India

EDITORIAL BOARDS

2014-	Editorial Board Member	Journal of Computational Systems Biology
2015-	Editorial Board Member	Journal of Cognitive Systems Research
2013-	Editorial Board Member	Springer Open Journal of Big Data
2012-16	Associate Editor	IEEE/ACM Transactions on Bioinformatics and Comp. Biol.
2007-	Editorial Board Member	Journal of Bioinformatics and Biology Insights
2007-	Editorial Board Member	Intl. J. Computational Biology and Drug Design
2007-	Editorial Board Member	Intl. J. Functional Informatics & Personalized Medicine
2007-08	Editorial Board Member	Applied Intelligence Journal
2006-	Editorial Board Member	Intl. J. Semantic Web and Information Systems
2006-	Editorial Board Member	Springer Book Series: Advanced Info. & Knowledge Proc.
2002-05	Editorial Board Member	Machine Learning Journal
1999-15	Co-Editor-in-Chief	Cognitive Systems Research
2005-08	Editorial Board Member	Intl. J. Data Mining and Bioinformatics
2004-	Editorial Board Member	Intl. J. Information and Computer Security
2001	Guest Editor	Machine Learning Journal

STUDY SECTION AND REVIEW PANEL MEMBERSHIP

2020	Proposal Review Panels	NSF
2020	Special Emphasis Panel	NIH
2019	Proposal Review Panels	NSF
2018	Proposal Review Panels	NSF
2017	Proposal Review Panels	NSF
2016	Proposal Review Panels	NSF
2016	SBE Proposal Review Panel	NSF
2016	Special Emphasis Panel, BD2K T32	NIH
2016	Special Emphasis Panel, Precision Health Cohorts	NIH
2016	NCATS Special Emphasis Panel	NIH
2016	NIAIDS Special Emphasis Panel	NIH

2016	Review Panel	Canada Res. Chairs
2015	NCATS Special Emphasis Panel	NIH
2015	NCI Special Emphasis Panel	NIH
2015	SBE Proposal Review Panel	NIH
2015	NIAIDS Special Emphasis Panel	NIH
2015	Special Emphasis Panel, BD2K T32	NIH
2014	CISE Proposal Review Panel	NSF
2014	NAGMS Council (ad hoc member)	NIH
2013	CISE Proposal Review Panel	NSF
2013	Proposal Review Panel	DOE
2010	Special Emphasis Study Section, Natl. Ctr. for Res. Resources	NIH
2010	National Centers for Biomedical Computing Review Panel	NIH
2010	Study Section, Biological Data Management and Analysis	NIH
2010	CISE Review Panels	NSF
2009	Bioinformatics Resource Centers for Infectious Diseases	NIH
2009	Genomics, Bioinformatics and Systems Biology Review Panel	European Res. Council
2009	Review Panel	Canada Res. Chairs
2009	CISE Review Panel	NSF
2008	CISE Review Panel	NSF
2008	Study Section Chair, Data Ontologies, Sharing Data & Tools	NIH
2004-07	Biological Data Management & Analysis	NIH
2006	CISE Review Panel	NSF
2005	CISE Review Panel	NSF
2004	Special Study Section, NIH Roadmap Initiative	NIH
2004	Special Study Section, Data Sharing Initiative	NIH
2001-04	Special Study Section, Bioinformatics	NIH
2003	CISE Intelligent Information Systems Review Panel	NSF
2003	CISE CAREER Review Panel	NSF
2003	SBIR Review Panel	NSF
2003	SBIR Review Panel	NSF
1999	IGERT Review Panel	NSF
2002	CISE Review Panel	NSF
1994	CISE Review Panel	NSF

PROFESSIONAL AFFILIATIONS

2014-	Senior Member, AAAI, Association for Advancement of Artificial Intelligence)
1986-2014	Member, AAAI (Association for Advancement of Artificial Intelligence)
2007-	Senior Member, ACM (Association for Computing Machinery)
2005-	Senior Member, IEEE (Institute of Electrical and Electronic Engineers)
1990-	Member, Institution of Electrical and Electronic Engineers (IEEE)
1990-2005	Member, IEEE Computer Society
1990-2007	Member, Association of Computing Machinery (ACM)
1992-	Member, American Association for Advancement of Science
1998-	Member, Cognitive Science Society
2003-	Member, International Society for Computational Biology

RESEARCH STATEMENT

My research interests cut across Computer and Information Science, Statistics, Biological Sciences, and Cognitive Sciences. This research is driven by fundamental scientific questions or important practical problems such as the following:

- How can we efficiently build predictive models from large, high-dimensional, distributed, autonomous, semantically disparate, data ("big data")?
- How can we elicit causal relations from multiple disparate observational and experimental studies?
- How can we ensure that AI systems in general and machine learning systems in particular are fair, explainable, and accountable?
- How can we extract useful knowledge from richly structured data (sequences, images, networks, etc.)?
- How can we build predictive models from multi-modal (multi-view) data?
- What are the information requirements and algorithmic basis of learning in specific scenarios?
- How can we learn language syntax (grammars) and semantics?
- How can we build predictive models from longitudinal data?
- How can computational abstractions of scientific artifacts and processes mediate scientific collaborations that transcend disciplinary and organizational boundaries and accelerate science?
- How can we efficiently represent and reason about preferences?
- How can we query and reason with federated data and knowledge bases?
- How can we assemble, adapt, and execute complex services from component services?
- How can we answer queries against knowledge bases without revealing secrets?
- How can we support data access and use policy compliant data integration and analysis?
- How can we learn to predict health outcomes and perform interventions from clinical, biomedical, environmental, socio-demographic, behavioral, and other types of data?
- How can we predict protein-protein, protein-RNA, protein-DNA interactions, interfaces, and complexes?
- How can we construct, compare and analyze multi-scale, models of molecular networks that orchestrate cellular development, differentiation, immune response, etc.?
- How can we model, construct, compare, and analyze brain networks from data?
- How can we build robust intelligent agents that incorporate multiple facets of intelligence?

My research contributions have spanned Computer Science (especially on the topics of Machine Learning and Data Mining, Knowledge Representation and Inference, Causal Inference) and in Bioinformatics and Computational Biology (especially on the topic of analysis and prediction of biomolecular (protein-protein, protein-DNA, and protein-RNA) interfaces and comparative analysis of biomolecular interaction networks). Some of my most recent work has focused on (i) Scalable algorithms for building predictive models from large, distributed, semantically disparate data (big data), including more recently, linked open data (ii) Algorithms for constructing predictive models from sequence, image, text, multi-relational, graph-structured data; (iii) New approaches to selective sharing of knowledge across autonomous knowledge bases (including knowledge base federation, secrecy-preserving query answering); (iv) Theoretically sound yet practically useful approaches to functional and non-functional specification driven composition of complex services from components;
(v) Expressive languages for representing, and model checking approaches to reasoning with,

qualitative preferences; (vi) Algorithms for eliciting causal effects from disparate sources of observational and experimental data; (vii) Scalable algorithms and software for comparative analyses of large bio-molecular networks and (viii) Machine learning approaches to analysis and prediction of macromolecular interactions and interfaces (including in particular, the first algorithm for partner- specific prediction of protein-protein interface sites and state-of-the-art sequence based protein-RNA interface predictors) that have resulted in several widely used web servers for analysis and prediction of protein-protein, protein-DNA, and protein-RNA interactions and interfaces, B-cell and T-cell epitopes.

Over the next 5-10 years, I plan to focus my research on (1) Computational abstractions scientific artifacts (e.g., data, knowledge, hypotheses), and universes of scientific discourse (e.g., biology), and scientific processes (e.g., hypothesis generation, predictive modeling, experimentation, simulation, and hypothesis testing), cognitive tools that augment and extend human intellect; and human- machine cyberinfrastructure (including organizational structures and processes) to accelerate science; (2) Design and analysis of algorithms for predictive modeling from very large, high dimensional, richly structured, multi-modal, longitudinal data; (3) Elucidation of causal relationships from disparate experimental and observational studies; (4) Elucidation of causal relationships from relational, temporal, and temporal-relational data; (5) Design and analyses of accountable, explainable, and fair AI systems; (6) Analysis and prediction of macromolecular interactions, elucidation of complex biological pathways e.g., those involved in immune response, development, and disease; (6) Predictive and causal modeling of individual and population health outcomes from behavioral, biomedical, clinical, environmental, socio- demographic data; (7) Predictive and causal modeling of behavioral and cognitive systems in naturalistic settings; (8) Modeling the structure, activity, and function of brain networks from fMRI and other types of data; (9) Predictive modeling of material properties from composition and adaptive design of materials; (10) Algorithmic fairness criteria and their operationalization.

Current Research Interests

1. **Artificial Intelligence:** Logical, probabilistic, causal and decision-theoretic knowledge representation and inference, Neural architectures for knowledge representation and inference, Computational models of perception and action. Intelligent agents and Multi-agent systems.
2. **Machine Learning, Data Mining, and Big Data Analytics:** Statistical, information theoretic, linguistic and structural approaches to machine learning, Learning and refinement of Bayesian networks, causal networks, decision networks, neural networks, support vector machines, kernel classifiers, multi-relational models, language models (n-grams, grammars, automata), grammars; Learning classifiers from attribute value taxonomies and partially specified data; Learning attribute value taxonomies from data; Learning classifiers from sequential and spatial data; Learning relationships from multi-modal data (e.g., text, images), Learning classifiers from distributed data, multiple instance data, multiple instance, multiple class data; networked data; multi-relational data, linked open data (RDF), and semantically heterogeneous data; Incremental learning, Ensemble methods, multi-agent learning, curriculum-based learning; selected topics in computational learning theory; Explainable machine learning; Knowledge guided machine learning.
3. **Bioinformatics and Computational Molecular and Systems Biology:** Data-driven discovery of macromolecular sequence-structure-function-interaction-expression relationships, identification of sequence and structural correlates of protein-protein, protein-RNA, and protein-DNA interactions, protein sub-cellular localization, automated protein structure and function annotation, modeling and inference of genetic regulatory networks from gene expression (micro-array, proteomics) data, modeling and inference of signal transduction and metabolic pathways, comparative analysis of biological networks (network alignment), integrative analysis of molecular interaction networks and macro-molecular interfaces.

4. **Discovery Informatics:** Computational models of scientific discovery; Discovery informatics infrastructure to integrate data, hypothesis, and knowledge-based inference, predictive modeling, experimentation, simulation, and hypothesis testing to provide an orderly formal framework and exploratory apparatus for science; Applications in computational systems biology.
5. **Knowledge Representation and Semantic Web:** Probabilistic, grammatical, network based, relational, logical, epistemic knowledge representation; knowledge-based, network based, and probabilistic approaches to information integration; description logics, federated data bases – statistical queries against federated databases, knowledge bases – federated reasoning, selective knowledge sharing, services – service composition, substitution, and adaptation; epistemic description logics; secrecy-preserving query answering, representing and reasoning about qualitative preferences, representing and reasoning about causality.
6. **Applied Information Integration and Informatics:** Applications of artificial intelligence, machine learning, and big data analytics to problems in bioinformatics, biomdical and health informatics, neuroinformatics, materials informatics, computational phenomics, security informatics, social informatics, and e-science.
7. **Other Topics of Interest:** Biological Computation – Evolutionary, Cellular and Neural Computation, Complex Adaptive Systems, Sensory systems and behavior evolution, Language evolution, Mimetic evolution; Computational Semiotics – Origins and use of signs, emergence of semantics; Computational organization theory; Computational Neuroscience; Computational models of creativity, Computational models of discovery.

RESEARCH, TRAINING, AND INFRASTRUCTURE GRANTS

Current Research, Training, and Infrastructure Grants[1]

1. Honavar, V. (PI). Interpreting Black-Box Predictive Models Through Causal Attribution. National Science Foundation. $200,000. 2020-2022.
2. Honavar, V. (PI). AI Institute: Planning: Institute for AI-Enabled Materials Discovery, Design, and Synthesis. National Science Foundation. $500,000. 2020-2022.
3. Honavar, V. (Co-PI). CC* CRIA: The Eastern Regional Network. National Science Foundation. $202,618. 2020-2022.
4. Honavar, V. (Co-PI). Scalable Near-Real-Time Identification and Characterization of Malware Behaviors Using Darknet Data. William J. Hughes Technical Center. $249,776. 2019-2021.
5. Honavar, V. (Co-PI). Graph Optimization via Deep Learning. Samsung Advanced Institute of Technology. $146,661. 2020-2021.
6. Honavar, V. (Co-PI), "Penn State Clinical and Translational Science Institute (UL1)," National Center for Advancing Translational Sciences, National Institutes of Health. $18,400,465. 2016 - 2020.
7. Honavar, V. (PI), "CIF21 DIBBs: EI: Virtual Data Collaboratory: A Regional Cyberinfrastructure for Collaborative Data Intensive Science," National Science Foundation, Collaborative Project with Rutgers University, Total budget: $4,000,000. Penn State Budget: $1,482,149. 2017 -2022.
8. Honavar, V. (Co-PI). BD Hubs: NORTHEAST: The Northeast Big Data Innovation Hub. National Science Foundation, $4,000,000. 2019-2023.
9. Honavar, V. (PI), "BD Spokes: SPOKE: NORTHEAST: Collaborative Research: Integration of environmental factors and causal reasoning approaches for large-scale observational health research," National Science Foundation, Collaborative Project with Harvard, Columbia, and CMU. Total budget: $1,000,000. Penn State Budget: $95,367. 2017 – 2020.
10. Honavar, V., Ritchie, M. D., Li, R. (Pis), "Biomedical Big Data to Knowledge (B2D2K) Training Program," National Institutes of Health, $1,604,971.00. 2016 – 2021.

[1] All the NSF grants that were active when I began my IPA appointment at NSF were assigned substitute PIs to manage them during my tenure at NSF. I could not serve as a PI or co-PI on grants submitted to NSF during 2010-2013.

Pending Grant Proposals

1. Collaborative Research: DMREF: Low-Dimensional Hybrid Organic–Inorganic Perovskites as Emergent Ferroelectric Materials for Next Generation of Optoelectronics, Vasant Honavar (Co-PI), National Science Foundation, $800,000. 2021-2024.
2. Mid-scale RI-1 (M1:IP): The Eastern Regional Network Federated Open Cyberinfrastructure (CI) Collaboratories (OpenCI Labs) for Research and Education, Vasant Honavar (Co-PI), National Science Foundation, $19,743,050. 2021-2024
3. FAI: Operationalizing Algorithmic Fairness Using Tools of Causal Inference, Vasant Honavar (PI), National Science Foundation, $1,155,085, 2021-2023.
4. Collaborative Research: RESEARCH-PGR: Genetic and environmentally-induced functional variation in the rice RNA structurome, Vasant Honavar (Co-PI), National Science Foundation, $2,028,882, 2021-2025.
5. Penn State Clinical and Translational Science Institute, Vasant Honavar (Co-PI), National Institutes of Health, $22,113,244. 2021-2026.
6. Towards Dynamic Patient-centric Personal Health Libraries, Vasant Honavar (Co-PI), National Institutes of Health, $1,510,407. 2021-2024. Under revision.

Past Research and Training Grants

1. Honavar, V. (PI), "Towards Computational Infrastructure for Analysis of Sensitive Data," National Science Foundation, $231,578.00. 2015 – 2018
2. Honavar, V. (PI), "SHF: Large: Collaborative Research: Inferring Software Specifications from Open Source Repositories by Leveraging Data and Collective Community Expertise,". National Science Foundation, Collaborative Project with a budget of $2.5 million, Penn State Budget: $319,511.00. 2015 – 2020.
3. Honavar, V. (PI), "Automating Artificial Intelligence: Empowering Analysts with Intelligent, Autonomous Software Agents," NGIA (Subaward from Concurrent Technologies Corporation), Federal Agencies. Total requested: $67,279., 2017.
4. New Techniques Towards Active Cyberdefense, National Security Agency, Vasant Honavar (Co-PI), $275,143.00. 2015 – 2016.
5. Identifying porcine genes and gene networks involved in effective response to PRRS virus using functional genomics and systems biology, Joan Lunney (PI), Vasant Honavar (Co-PI), Zhihua Jiang (Co-PI), Roman Pogranichniy (Co-PI), Juan Pedro Steibel (Co-PI), Chris Tuggle (Co-PI), United States Department of Agriculture, 2010-2014, $750,000.
6. Honavar, V. (Mentor), Coffman, D. L., Li, R., Fellowship, "Novel Methods to Identify Momentary Risk States for Stress and Physical Inactivity," National Institutes of Health, $500,000.00. 2014- 2017).
7. Collaborative Research: Learning Classifiers from Autonomous, Semantically Heterogeneous Distributed Data Sources, National Science Foundation, Vasant Honavar (PI) $527,000 (including supplements). 2007-2013.
8. Intergovernmental Personnel Act (IPA) Appointment, Program Director, Information Integration and Informatics Program, Information and Intelligent Systems Division, Computer and Information Science and Engineering Directorate, National Science Foundation, $701,972. 2010-2013
9. IGERT: Computational Molecular Biology Training Program. Vasant Honavar (Co-PI) with Dan Voytas (PI), Drena Dobbs (Co-Pis). National Science Foundation, 2005-2012. $2,968,976.

10. High-Accuracy Protein Models Derived from Lower Resolution Data. National Institutes of Health (2007-2010), Vasant Honavar (Co-PI), with Andrzej Kloczkowski (PI), Robert Jernigan, Mark Gordon, Zhijun Wu, Iowa State University and Janusz Bujnicki, Krzysztof Ginalski and Andrzej Kolinski, Warsaw University (Co-Pis), $744,725.
11. Interactive and Verifiable Composition of Web Services To Satisfy End User Goals. National Science Foundation, Vasant Honavar (Co-PI), with Samik Basu (PI) and Robyn Lutz (Co-PI). (2007-2011), $350,002.
12. Machine Learning Algorithms and Software for Collaborative Medical Decision Support. Grow Iowa Values Fund, and Collaborative Health Solutions, LLC 2010-2012, $238,586
13. Developing Predictive Models for Identifying Pigs with Superior Immune Response and Improved Food Safety, Chris Tuggle (PI), Bearson, S., Honavar, V., Nettleton, D. Wannemuehler, M., Lunney, J. and Nettleton, D. (Co-Pis). United States Department of Agriculture (2009-2012), $1,000,000.
14. NIH-NSF BBSI Summer Institute in Bioinformatics and Computational Biology – Iowa State University. National Science Foundation, Vasant Honavar (Co-PI) with Volker Brendel (PI), Robert Jernigan, Karin Dorman, and Julie Dickerson (Co-Pis) (2006–2010). $499,000.
15. Development of bioinformatics resources to transfer biological information across species. *United States Department of Agriculture.* Vasant Honavar (Co-PI), James Reecy (PI), Anne Kwitek (Co-PI). (2008-2010). $1,000,000.
16. DDAS-TMRP: Auto-Steered Information-Decision Processes for Electric Power Systems Asset Management. *National Science Foundation,* Vasant Honavar (Co-PI) with James McCalley (PI), Sarah Ryan (Co-PI), William Meeker (Co-PI), and Daji Qiao (Co-PI). (2006-2011) $700,000.
17. Discovering Protein Sequence-Structure-Function Relationships*.* National Institutes of Health Vasant Honavar (PI), Robert Jernigan and Drena Dobbs (Co-Pis), (2003-2008). $1,022,000.
18. IIS: Exploratory Investigation of Modular Ontologies. *National Science Foundation,* Vasant Honavar (PI), Giora Slutzki and Doina Caragea (Co-Pis), (2006-2008). $112,000.
19. Integration of Functional Genomics and Quantitative Genetics to Improve Feed Efficiency in Pigs. United States Department of Agriculture (2005-2008), Vasant Honavar (Co-PI) with Jack Dekkers (PI), Chris Tuggle (Co-PI), Dan Nettleton (Co-PI), Lloyd Anderson (Co-PI), Rondhane Rekaya (Co-PI), Richard Barb (Co-PI), and Gary Hausman (Co-PI), $876,000.
20. Computational Support Staff for Expanding Animal Functional Genomics Capabilities. Vasant Honavar (PI), Chris Tuggle, Jim Reecy, Diane Spurlock, Jack Dekkers, Susan Lamont, Chad Stahl (Co-Pis), $50,000. (2007-2009).
21. Center for Computational Intelligence, Learning, and Discovery. Vasant Honavar (PI). Vice Provost for Research, Iowa State University (2004-2009) $477,500
22. ITR: Algorithms and Software for Knowledge Acquisition from Heterogeneous Distributed Data. *National Science Foundation*, Vasant Honavar (PI) Drena Dobbs (Co-PI), (2002-2007). $223,500.
23. Algorithms and Software for Collaborative Ontology Development. Center for Integrated Animal Genomics, Iowa State University. Vasant Honavar (PI), (2005-2007) $25,000.
24. IGERT: Computational Molecular Biology Training Program. Vasant Honavar (Co-PI) with Dan Voytas (PI), Pat Schnable, Susan Carpenter, Jonathan Wendel (Co-Pis). National Science Foundation*,* 1999-2004., $2,374,597 (plus $1,161,010 in matching funds).
25. Developmental Proteomics of Retinal Progenitor Cells, National Institutes of Health (2003-2006). Vasant Honavar (Co-PI), (with Heather West-Greenlee and Jan Buss), $438,000.

26. NIH-NSF BBSI Summer Institute in Bioinformatics and Computational Biology – Iowa State University. National Science Foundation, Vasant Honavar (Co-PI) with Volker Brendel (PI), Robert Jernigan, Karin Dorman, and Xun Gu (Co-Pis) (2002-2006). $645,000.
27. Automated Integration of Condition Monitoring with an Optimized Maintenance Scheduler for Circuit Breakers and Power Transformers. Vasant Honavar (Co-PI) with James McCalley (PI) Mladen Kezunovic, and Chanan Singh (Co-Pis), Power Systems Engineering Research Center (a National Science Foundation Industry-University Research Center), 2002-2005. (2002-2005). $255,000.
28. Symposium on Integration of Structural and functional Genomics, Vasant Honavar (co-PI) with Chris Tuggle (PI) and Marit Nielsen-Hamilton (Co-PI) National Science Foundation. (2005). $10,900.
29. Exploring a Novel Proline Switch for Regulation of Protein Recognition. Center for Integrated Animal Genomics, Iowa State University, Vasant Honavar (Co-PI) with Drena Dobbs and Susan Carpenter (Co-Pis) (2004-2005) $30,000.
30. Constructive Neural Network Learning Algorithms for Pattern Classification, National Science Foundation, 1994-1999, Principal Investigator, $111,537 (plus $10,000 in matching funds).
31. SGER: Multidisciplinary Aspects of Computation Theory, National Science Foundation, Vasant Honavar (Co-PI), with Jack Lutz (PI), Pavan Aduri (Co-PI), and Krishna Athreya (Co-PI), (2003-2005). $74,948.
32. Graduate Research Fellowships in Bioinformatics and Computational Biology, Pioneer Hi-Bred, Inc. 2002-2004. Major professor, Adrian Silvesu and Carson Andorf, $80,000.
33. IBM Graduate Research Fellowship in Computer Science, IBM Inc., Major Professor, Doina Caragea, 2003-2004, $30,000 (17pprox..)
34. Interactive Visual Overviews of Large, Multi-Dimensional Datasets, Vasant Honavar (Co-PI) with Diane Cook (PI) and Les Miller (Co-PI) National Science Foundation, Co-Principal Investigator, 1999-2003, $370,000.
35. Innovative Technologies for Defense Against Catastrophic Failures of Complex, Interactive Power Networks, U.S. Department of Defense (DOD) and Electric Power Research Institute (EPRI), $4,500,000. (1999 – 2004) (Collaborative project involving 9 ISU faculty and faculty from 3 other universities). Co-Principal Investigator.
36. An Agent-Based System for Integration and Analysis of Distributed, Heterogeneous Plant Genome Databases. Pioneer Hi-Bred International, Inc., 2000-2002, Principal Investigator, $40,000.
37. IBM Graduate Research Fellowship in Computer Science, IBM Inc., Vasant Honavar (Major Professor), with Doina Caragea (doctoral student), 2002-2003, $30,000 (17pprox..)
38. Intelligent Multi-Agent Systems for Intrusion Detection, National Security Agency, 1998-2000, Co-Principal Investigator, $199,769.
39. Data Mining of Electric Power Usage Data to Develop Customer Profiles. Cooperative Research Proposal. Power Domain, Inc. (2001-2002). Vasant Honavar. $43,639.
40. SGER: Distributed Knowledge Networks to Support Security-Economy Decisions in Stressed Electric Power Systems. National Science Foundation, 2000-2001, Co-Principal Investigator, $99,999.
41. Artificial Intelligence Applications to Power System Management and Control, Electric Power Research Institute, Co-principal investigator. 1998-2000, $151,000
42. Distributed Knowledge Networks, John Deere Foundation, 1999-2001, Principal Investigator, $30,000.
43. Development of Algorithmic Approaches to Gene Expression Analysis from Microarray Data, *Carver Foundation*, 2000-2001, Principal Investigator, $25,000.

44. Development of Protein Structure Prediction Algorithms. Carver Foundation, 1999-2000, Co- Principal Investigator, $25,000.
45. Genetic algorithms for protein structure prediction. Ames Laboratory, Co-Principal Investigator, 1999-2000. $35,700.
46. A Gene-Specific DNA Chip for Exploring Molecular Evolutionary Change, *Carver Foundation*, 1998-1999. Co-Principal Investigator, $17,120.
47. Intelligent Diagnosis Systems, John Deere Foundation, 1995-1998, Principal Investigator, $30,000.
48. Graduate Fellowship (Data Mining and Knowledge Discovery), IBM Corporation, 1997-1998, Principal Investigator, $20,800.

SELECTED RESEARCH PROJECTS AND RESEARCH ACCOMPLISHMENTS

Predictive models from ultra-high-dimensional longitudinal data (supported in part by Frymoyer Chair in IST held by Vasant Honavar at Penn State University and NIH Clinical and Translational Sciences Institute)

Longitudinal data, sometimes also called panel data, i.e, collections of repeated observations from a set of individuals, taken from a larger population, over a period of time, often at irregularly spaced time points for each individual, are common across a broad range of applications, including health sciences, social sciences, learning sciences, economics, among others. Such data can be used to uncover the relationship between the time-varying patterns of certain measured variables (or features) and a particular outcome variable (or outcome) of interest, e.g., stock market crash, disease onset, health risk.

Longitudinal data exhibit longitudinal correlation (LC), i.e., correlations across observations of the same individual taken at different time points. In addition, observations across individuals may be correlated as well because of their shared traits (e.g., demographic characteristics), leading to clustered correlation (CC), or both. Under such circumstances, the observations (either within or across individuals) are no longer independent and identically distributed (i.i.d). Ignoring either part of the correlation can lead to incorrect parameter estimation, invalid tests of hypothesis, and misleading statistical inferences or predictions. Moreover, such data are characterized by fixed effects that are shared by the population under study; random effects that are individual-specific, or mixed effects, i.e., the combination of fixed effects and random effects. With the advent of big data, often the number of variables far exceeds the number of individuals, which greatly increases the need for effective variable selection, computational efficiency and interpretable models. Last, but not the least, observations at any given time point have many missing measurements, and missing data are generally not missing at random.

Against this background, we are developing novel methods for predictive modeling from ultra-high-dimensional, irregularly sampled, longitudinal data, including:

- Longitudinal Multi-Level Factorization Machines (LMLFM) (Liang et al., 2020), a novel, efficient, provably convergent extension of Factorization Machine (FM) for predictive modeling of longitudinal data characterized by mixed effects, in the presence of LC, CC, or both. A key feature of FM is that it models interactions of variables by mapping the interactions to dot products of vectors in a low dimensional latent space. LMLFM, like FM, uses latent factors to efficiently model higher order interactions between features. LMLFM extends FM to handle fixed, random, or mixed effects as needed for predictive modeling from longitudinal data. In addition, LMLFM incorporates a change in the structure of the underlying model to enhance model interpretability and achieve strictly linear training time with respect to the size of training data. To the best of our knowledge, LMLFM is the first multi-level regression model that extends variable selection beyond fixed effects to include random effects, and hence can be applied to high-dimensional longitudinal data. LMLFM

uses a hierarchical probabilistic graphical model (HPGM) and avoids the need for hyperparameter tuning. LMLFM uses a variant of the iterated conditional modes (ICM) algorithm for learning the parameters of an LMLFM model based on a maximum a posteriori (MAP) formulation derived from the HPGM. We have established the convergence of LMLFM. Results of experiments with simulated data that show that LMLFM can effectively cope with high dimensional longitudinal data in the presence of both LC and CC whereas state-of-the-art baseline methods fail to do so (e.g., LMLFM can handle longitudinal data with over 5000 variables whereas the state-of-the-art baseline multilevel mixed effects models fail when the number of variables exceeds 100). Results of experiments with real-world data sets show that LMLFM compares favorably with the state-of-the-art baselines in terms of predictive accuracy, while producing sparse, interpretable models that include only the relevant subset of variables.

- Longitudinal deep kernel Gaussian process regression (L-DKGPR) (Liang et al., 2021) to overcome these limitations by fully automating the discovery of complex multilevel correlation structure from longitudinal data. Specifically, L-DKGPR eliminates the need for ad hoc heuristics or trial and error using a novel adaptation of deep kernel learning that combines the expressive power of deep neural networks with the flexibility of non-parametric kernel methods. L-DKGPR effectively learns the multilevel correlation with a novel additive kernel that simultaneously accommodates both time-varying and the time-invariant effects. We have developed an efficient algorithm to train L-DKGPR using latent space inducing points and variational inference. Results of extensive experiments on several benchmark data sets demonstrate that L-DKGPR significantly outperforms the state-of-the-art longitudinal data analysis (LDA) methods.
- A novel, modular, convolution-based feature extraction and attention mechanism that simultaneously identifies the variables as well as time intervals over which the variables identified impact the classifier output (Hsieh et al., 2021). The results of our extensive experiments with several benchmark data sets that show that the proposed method outperforms the state-of-the-art baseline methods on multi-variate time series classification task. The results of our case studies demonstrate that the variables and time intervals identified by the proposed method make sense relative to available domain knowledge.
- Functional autoencoders (Hsieh et al., 2021), which generalize neural network autoencoders so as to learn non-linear representations of functional data. We have derived from first principles, a functional gradient based algorithm for training functional autoencoders. The results of experiments demonstrate that the functional autoencoders outperform the state-of-the-art baseline methods. The resulting methods find applications in many real-world settings, e.g., monitoring of individual health, climate, brain activity, environmental exposures, among others, the data of interest change smoothly over a continuum, e.g., time, yielding multi-dimensional functional data. Functional data representations can be used for functional data clustering, classification, and regression.
- SRVAR (Hsieh et al., 2021), a novel approach to the problem of simultaneously learning the dynamics of transitions between hidden states and the state-dependent relationships between variables. SRVAR uses state space recurrent neural networks to model the transitions between hidden states, and utilizes the smooth acyclic characterization of DAGs to efficiently learn the state-dependent DAGs. We have results of experiments on simulated data as well as a real-world data set that show the superiority of SRVAR over state-of-the-art baselines in recovering the patterns of state transitions while modeling state-specific dependencies between variables.

Predictive and Causal Modeling of Health Outcomes from Clinical, Behavioral, Biomedical, Socio-Demographic, and Environmental Data (Funded in part by grants from the National Science Foundation and the National Institutes of Health)

There is increasing recognition that environmental and contextual factors can have a significant impact on the health outcomes in diseases such as cancer, obesity, diabetes, heart disease. The advent of "big data" offers enormous potential for understanding and predicting health risks, intervention outcomes, and personalized treatments, ultimately improving population health through integrative analysis of heterogeneous, fine-grained, richly structured, longitudinal patient data. This project aims to bring together an interdisciplinary team of researchers to to understand the clinical, behavioral, biomedical environmental, and contextual (e.g., socio-demographic) factors that contribute to increased risk of specific diseases, e.g., breast cancer; and developing evidence-based practices for supplemental screening, as well as behavioral or clinical interventions to mitigate the risk. The project leverages the infrastructure of the PCORI-funded PaTH Clinical Data Research Network (CDRN), a consortium of four Mid-Atlantic academic health systems, including the Penn State College of Medicine Milton S. Hershey Medical Center (www.pathnetwork.org), for EHR data coded using standardized vocabularies. The key methodological and informatics innovations in the project have to do with the development of novel algorithms and tools for predictive modeling of health risks and health outcomes by integrating clinical, biomedical, environmental, socio-demographic and behavioral data. Work in progress is aimed at:

- Developing customizable, auditable, modular, data access and use policy compliant software workflows for integration of electronic health records (EHR) data with selected environmental, behavioral, biomedical, socio-demographic data
- Applying the workflow to securely assemble and share data sets that can be used to address specific clinical or biomedical, or population health related research questions;
- Develop novel algorithms for predictive and causal modeling of health risks from the resulting "big data"
- Elicit the environmental, contextual, behavioral, health status and health care factors that are reliable predictors of health risks, health outcomes, or effective interventions through integrative analysis of clinical, environmental, behavioral, biomedical, socio-demographic data.

Causal Inference (supported in part by Frymoyer Chair in IST held by Vasant Honavar at Penn State University)

Elicitation of a causal effect from observations and experiments is central to scientific discovery, or more generally, rational approaches to understanding and interacting with the world around us. Judea Pearl introduced causal diagrams provide a formal representation for combining data with causal information and do-calculus to provide a sound and complete inferential machinery for causal inference.

The practical need to transfer causal effects elicited in one domain (setting, environment, population) e.g., a controlled laboratory setting, to a different setting presents us with the problem of transporting causal information from a source environment to a possibly different target environment. For example, one might want to know if causal relation between teaching strategies and student learning obtained by through a randomized trial in a public school in Chicago can be transported to a public school in Minneapolis that has an admittedly different population of students. Our recent work has led to theoretical characterization and algorithms for:

(a) *m*-transportability, a generalization of transportability, which offers a license to use causal information elicited from experiments and observations in *m* (where *m* is greater than or equal to 1) source settings to estimate a causal effect in a given target setting. We have established the necessary and sufficient conditions for *m*-transportability. We have designed an algorithm for deciding *m*-transportability that determines whether a causal relation is m-transportable; and if it is, produces a transport formula, that is, a recipe for estimating the desired causal effect by combining experimental information from *m* source environments with observational information from the target environment (Lee et al., 2013)

(b) *z*-transportability, the problem of estimating the causal effect of a set of variables X on another set

of variables Y in a target domain from experiments on any subset of controllable variables Z where Z is an arbitrary subset of observable variables V in a source domain. Z-Transportability generalizes *z*-identifiability, the problem of estimating in a given domain the causal effect of X on Y from surrogate experiments on a set of variables Z such that Z is disjoint from X. z- Transportability also generalizes transportability which requires that the causal effect of X on Y in the target domain be estimable from experiments on any subset of all observable variables in the source domain. We have generalized z-identifiability to allow cases where Z is not necessarily disjoint from X. We have established a necessary and sufficient condition for z-transportability in terms of generalized z-identifiability and transportability. We have provided a sound and complete algorithm that determines whether a causal effect is z-transportable; and if it is, produces a transport formula, that is, a recipe for estimating the causal effect of X on Y in the target domain using information elicited from the results of experimental manipulations of Z in the source domain and observational data from the target domain. Our results also show that do-calculus is complete for z-transportability (Lee et al., 2013).

(c) *mz*-transportability, the problem of inferring a causal effect of treatment variables on observables in a target domain (environment, experimental setting) by combining data from experiments on simultaneously controllable subsets of variables (together with observations) from multiple domains (including the target domain). We have provided an efficient and complete algorithm that determines if a causal effect is mz-transportable, and if so, outputs a transport formula for estimating the causal effect. These results set the stage for considering more general forms of meta-identifiability by allowing a fully arbitrary information set and for proving the completeness of do-calculus in such settings (Bareinboim et al., 2013).

We have examined the following problem in causal inference: Given a causal graph G, determine MIC(G), that suffices for identifying every causal effect that is identifiable in a causal model characterized by G. We have established the completeness of do-calculus for computing MIC(G) . MIC(G) effectively offers an efficient compilation of all of the information obtainable from all possible interventions in a causal model characterized by G. Minimum intervention cover finds applications in a variety of contexts including counterfactual inference, and generalizing causal effects across experimental settings. We analyze the computational complexity of minimum intervention cover and identify some special cases of practical interest in which MIC(G) can be computed in time that is polynomial in the size of G.

Work in progress is aimed at (i) bridging the gap between theory and practice of causal inference to address the needs of real-world applications, e.g., by developing algorithms and software for eliciting causal effects from temporal and temporal-relational data.

Most work on causal inference from observational data has assumed that the observations are independent and identically distributed. However, in many practical applications, the data exhibit relational dependencies. Relational causal models (RCM) allow us to model causal relationships in relational data. Our recent work has led to:

(a) A characterization of the properties of abstract ground graphs (AGG), which play a key role in the proofs of completeness of the only previous algorithm for learning relational causal models from data, We showed that AGG representation is not complete for relational *d*-separation, that is, there can exist conditional independence relations in an RCM that are not entailed by AGG (Lee et al., 2015).

(b) An investigation of Relational Causal Models (RCM) under relational counterparts of adjacency-faithfulness and orientation-faithfulness, yielding a simple approach to identifying a subset of relational *d*-separation queries needed for determining the structure of an RCM using *d*-separation against an unrolled DAG representation of the RCM. We provided theoretical underpinnings of a basis of a sound and efficient algorithm for learning the structure of an RCM

from relational data. We introduced RCD-Light, a sound and efficient constraint-based algorithm that is guaranteed to yield a correct partially-directed RCM structure with at least as many edges oriented as in that produced by RCD, the only other existing algorithm for learning RCM. We showed that unlike RCD, which requires exponential time and space, RCD- Light requires only polynomial time and space to orient the dependencies of a sparse RCM (Lee et al., 2016).

(c) A novel and elegant characterization of the Markov equivalence of RCMs under *path semantics, an alternative to bridge-burning semantics used by RCD.* We introduced a novel representation that allows us to efficiently determine whether an RCM is Markov equivalent to an-other. Under path semantics, we provide a sound and complete algorithm for recovering the structure of an RCM from conditional independence queries. Our analysis also suggests ways to improve the orientation recall of algorithms for learning the structure of RCM under *bridge burning semantics* as well (Lee et al., 2016).

Federated Data and Computational Infrastructure for Collaborative, Data-Intensive Science
(funded in part by a grant from the National Science Foundation)

Scientific progress in many disciplines is increasingly enabled by our ability to examine natural phenomena through the computational lens (e.g., using algorithmic abstractions of the underlying processes) and our ability to acquire, share, integrate, and analyze disparate types of data. However, realizing the full potential of data to accelerate science calls for significant advances in data and computational infrastructure to support collaborative data-intensive science by teams of researchers that transcend institutional and disciplinary boundaries.

This project aims to conceptualize, design, and implement a Virtual Data Collaboratory (VDC), to support collaborative, data-intensive science research by multi-disciplinary teams drawn from multiple institutions. Specifically, the project aims to design VDC, a federated infrastructure that integrates the state of the art data-intensive computing platforms, storage, and networking, with an innovative data services layer across Rutgers University, Pennsylvania State University, and several other institutions in the region, interconnected through a high-speed network, with the potential to expand to incorporate academic/research institutions across the United States. VDC will leverage existing national/international and regional data repositories (including NSF funded repositories like the Ocean Observatories Initiative (OOI) and the Protein Data Bank (PDB)), existing investments in advanced cyberinfrastructure, like the NSF funded Big Data Regional Hubs, XSEDE, OSG, among others.

VDC will provide the collaborative infrastructure and platform for developing and integrating algorithmic abstractions of scientific domains e.g., biology, coupled with methods and tools for data analytics, modeling, and simulation, cognitive tools (representations, processes, protocols, workflows, software) to advance science. VDC will support reproducible, sharable, and reconfigurable data- intensive scientific workflows [Parashar et al., 2019].

The project will use several collaborative science use cases to develop and evaluate the VDC infrastructure. For example, one use case involves a collaboration between Vasant Honavar and Helen Berman, a Rutgers structural biologist and the founder of the Nucleic Acid Database (NDB) and former director of the Protein Data Bank (PDB), a widely used archival database of curated protein structures, will use VDC to assemble carefully curated data sets of protein-DNA and protein RNA complexes and interfaces; and develop machine learning and other computational methods and tools for reliable prediction of protein-RNA and protein-DNA interfaces. The team will use VDC to establish shared data and computational infrastructure, complete with workflows for documenting, comparing, and reproducing computational analyses and prediction of protein-RNA complexes, interfaces. In addition to helping develop and evaluate the VDC infrastructure, the results of this effort will advance our understanding of the molecular mechanisms by which proteins recognize and bind to DNA and RNA, and

their role in a variety of important biological processes that orchestrate development, aging, disease, etc.

Algorithms and Software for Knowledge Acquisition from Semantically Heterogeneous, Distributed Data (funded in part by grants from the National Science Foundation)

The emergence of big data offers unprecedented opportunities in knowledge acquisition (e.g., discovery of a priori unknown complex relationships, construction of predictive models) from data. However, realizing these opportunities in practice presents several challenges. Data repositories are autonomously designed and operated, large in size, physically distributed, and differ in structure, organization, semantics, and query and processing capabilities. Our research, aimed at addressing some of these challenges, has led to:

(a) The development of a general theoretical framework for learning predictive models (e.g., classifiers) from large, physically distributed data sources where it is neither desirable nor feasible to gather all of the data in a centralized location for analysis. This framework (Caragea et al., 2001; 2003; 2004a) offers a general recipe for the design of algorithms for learning from distributed data that are *provably exact* with respect to their centralized counterparts (in the sense that the model constructed from a collection of physically distributed data sets is provably identical to that obtained in the setting where the learning algorithm has access to the entire data set). A key feature of our approach is the clear separation of concerns between hypothesis construction and extraction and refinement of *sufficient statistics* needed by the learning algorithm from data which reduces the problem of learning from data to a problem of decomposing a query for sufficient statistics across multiple data sources and combining the answers returned by the data sources to obtain the answer for the original query. Our work has resulted in provably exact algorithms (relative to the centralized counterparts) for learning a variety of predictive models from distributed data.

(b) The development of theoretically sound yet practical variants of a large class of algorithms (Caragea et al., 2001; 2003; 2004a; Koul et al., 2010; Lin et al., 2011; 2013) for learning predictive models (classifiers) from distributed data sources under a variety of assumptions (motivated by practical applications) concerning the nature of data fragmentation, and the query capabilities and operations permitted by the data sources (e.g., execution of user supplied procedures), and precise characterization of the complexity (computation, memory, and communication requirements) of the resulting algorithms relative to their centralized counterparts.

(c) The development of a theoretically sound approach to formulation and execution of statistical queries across semantically heterogeneous data sources (Caragea et al., 2004b; Caragea et al., 2005; Caragea et al., 2006; 2007; 2010). This work has shown how to use semantic correspondences and *mappings* specified by users from a set of terms and relationships among terms (user ontology) to terms and relations in data source specific ontologies to construct a sound procedure for answering queries for *sufficient statistics* needed for learning classifiers from semantically heterogeneous data. An important component of this work has to do with the development of statistically sound approaches to handling data specified at different levels of abstraction across different data sources (Zhang et al., 2003, 2006).

(d) Abstraction-Driven Algorithms for Building Compact yet Accurate Classifiers. We have developed a general approach for exploiting attribute value hierarchies (AVH) that group the values of attributes to learn compact yet accurate predictive models from data specified at different levels of abstraction. Instantiations of this approach in the case of Naïve Bayes (Zhang et al., 2004; 2006), Decision Trees (Zhang et al., 2003), and Markov Models (Caragea et al., 2010) show that the resulting algorithms yield predictive models that are more compact than those produced by their counterparts that do

not have access to AVH without sacrificing the quality of the predictors.

(e) Demonstration of the theoretical equivalence of a certain class of inter-ontology mapping errors and noise models, and hence the reduction of the problem of learning in the presence of mapping errors from semantically disparate data to the problem of learning from noisy data (Koul et al., 2010, 2012).

(f) The design and implementation of INDUS – A modular, extensible, open source software toolkit for data-driven knowledge acquisition from large, distributed, autonomous, semantically heterogeneous data sources (http://code.google.com/p/induslearningframework/; http://code.google.com/p/indusintegrationframework/).

Research in progress is aimed at:

(a) Extension of the statistical query based learning framework to learning predictive models from Linked Open (RDF) Data e.g., algorithms for learning Relational Bayesian Classifiers from RDF data in settings where the learner can access the RDF data only through a restricted set of queries against an access interface (Lin et al., 2011).

(b) Extension of the statistical query based learning framework to learning predictive models from network data

Applications of the resulting algorithms to social network and social media analytics and analysis and prediction of biomolecular interaction.

Learning Predictive Models from Richly Structured Data

Learning Predictive Models from Partially Specified, Sparse Data: Many applications, e.g., medical diagnosis, different cases may be described in terms of symptoms or results of diagnostic tests are only partially specified e.g., a patient may be described as having cardiac arrhythmia without specifying the precise type of arrhythmia. Other applications present the problem of learning predictive models from sparse data, e.g., each patient exhibits only a small subset of possible symptoms. My students and I pioneered a principled approach to learning compact yet accurate predictive models from data specified at varying levels of precision or abstraction and demonstrated how the approach could be used to learn accurate, and compact (and hence easy-to-comprehend) predictive models from sparse data, noisy data. The resulting methods have found applications in bioinformatics, health informatics, and social informatics.

Some results of this work include:

(a) Abstraction Augmented Markov models (AAMMs). AAMMs are generalizations of Markov Models (MM). AAMMs simplify the data representation used by the standard MMs by grouping similarsubsequences to organize them in an abstraction hierarchy (Caragea et al., 2010a, 2010b, 2010c). Experimental results on text document classification and protein subcellular localization show that adapting data representation by combining super-structuring and abstraction makes it possible to construct predictive models that use substantially smaller number of features (by one to three orders of magnitude) than those obtained using super-structuring alone (whose size grows exponentially with the length of direct dependencies). Super-structuring and abstraction-based models are competitive with and, in some cases, outperform, models that use only super- structuring. Our experiments have also demonstrated the promise of abstraction-augmented Markov Models in learning sequence classifiers in a semi-supervised setting where only some of the sequences are labeled.

(b) Development of Abstraction-super-structuring Normal Forms (Silvescu and Honavar, 2011) that offer a general theoretical framework for structural (as opposed to parametric) aspects of induction using abstraction (grouping of similar entities) and super-structuring (combining topologically close entities) and exploration of its relation to ideas e.g., radical positivism in the philosophy of science (with PhD student Adrian Silvescu).

(c) Development of methods for taking advantage of abstractions to build compact yet accurate predictive models for labeling actors in social networks (Bui et. al., 2013) and for building predictive models RDF data by taking advantage of RDFS hierarchies (Lin et al., 2015).

Work in progress is aimed at extending this approach to learning predictive models from richly structured data at multiple levels of abstraction (images and text (multi-modal data), social networks and social media, linked open data, biomolecular interaction network data).

Learning Predictive Models from Multi-Modal, Multi-View Data

Recent years have witnessed rapid advances in our ability to acquire and store massive amounts of data across different modalities (such as text, speech, images, etc. on the web; genomics, transcriptomics, proteomics, metabolomics data in life sciences; fMRI, diffusion tensor imaging, etc. in brain sciences; Different types of digital objects e.g., movies, music, etc. associated with individuals in a social network). Some of this work also had to cope with the challenges presented by the limited availability of labeled data, and the need for accommodating multiple labels for data instances. We have developed multimodal machine learning algorithms for image annotation, the task of assigning keywords to an image based on its contents, labeling actors in social networks (e.g., with respect to their interest in specific content, e.g., movies, music, etc). Some results of this work include:

(a) Multimodal Hierarchical Dirichlet Process Models, non-parametric generalizations of the hierarchical mixture models. Our experimental evaluation shows that the performance of this model does not depend on the number of mixture components, unlike the standard mixture model which suffers from over-fitting (Yakhnenko and Honavar, 2009).

(b) Formulation of image annotation as a multiple instance, multiple label learning problem, which is a generalization of supervised learning in which the training examples are bags of instances and each bag is labeled with a set of labels. We learn as many classifiers as there are possible labels and force the classifiers to share weights using trace-norm regularization. Our experimental results on standard benchmark datasets show that the performance of this model is comparable to the state- of-the-art multiple instance multiple label classifiers and that unlike some state-of-the-art models, it is scalable and practical for datasets with a large number of training instances and possible labels (Yakhnenko and Honavar, 2011).

(c) A generalization of the discriminative model to a semi-supervised setting to allow the model take advantage of labeled and unlabeled data. We assume that the data lies in a low-dimensional manifold and introduce a penalty that ensures that the classifiers assign similar labels to similar instances (i.e. instances that are near-by in the manifold induced by the samples). Our experimental results show the effectiveness of this approach in learning to annotate images from partially labeled data (Yakhnenko and Honavar, 2010).

(d) A novel kernel based approach for predicting labels of actors in multi-view social networks. Here, given a social network in which only a subset of the actors are labeled, our goal is to predict the labels of the rest of the actors. We introduced a new random walk kernel, namely the Inter-Graph Random Walk Kernel (IRWK), for labeling actors in multi-view social networks. IRWK combines information from within each of the views as well as the links across different views that outperform or are competitive with several state-of-the-art methods for labeling actors in social networks (Bui et al., 2016).

(e) A novel use of knowledge graphs, that capture general or commonsense knowledge, to augment the information extracted from images by the state-of-the-art methods for image captioning. The results of our experiments, on several benchmark data sets such as MS COCO, as measured by CIDEr-D, a performance metric for image captioning, show that the variants of the state-of-the- art methods for image captioning that make use of the information extracted from knowledge graphs can substantially outperform those that rely solely on the information extracted from images (Zhou et al., 2019).

(f) A novel multi-view network embedding (MVNE) algorithm for constructing low-dimensional node embeddings from multi-view networks. MVNE adapts and extends an approach to single view network embedding (SVNE) using graph factorization clustering (GFC) to the multi-view setting using an objective function that maximizes the agreement between views based on both the local and global structure of the underlying multi-view graph. Our experiments with several benchmark real-world single view networks show that GFC-based SVNE yields network embeddings that are competitive with or superior to those produced by the state-of-the-art single view network embedding methods when the embeddings are used for labeling unlabeled nodes in the networks. Our experiments with

several multi-view networks show that MVNE substantially outperforms the single view methods on integrated view and the state-of-the-art multi-view methods. We further show that even when the goal is to predict labels of nodes within a single target view, MVNE outperforms its single-view counterpart suggesting that the MVNE is able to extract the information that is useful for labeling nodes in the target view from the all of the views (Sun et al., 2018).

(g) A novel optimization algorithm for y machine learning, and statistical inference problems that call for minimization of a composition of expected value functions (CEVF). We have introduced compositional stochastic average gradient descent (C-SAG) a novel extension of the stochastic average gradient method (SAG) to minimize composition of finite-sum versions of such compositional optimization problems. C-SAG, like SAG, estimates gradient by incorporating memory of previous gradient information. We have presented detailed theoretical analyses of C- SAG which show that C-SAG, like C-SVRG, achieves a linear convergence rate for strongly convex objective function; However, C-CAG achieves lower oracle query complexity per iteration than C- SVRG. Results of experiments showing that C-SAG converges substantially faster than full gradient (FG), as well as C-SVRG (Hsieh et al., 2018).

Work in progress is aimed at developing efficient methods for learning accurate predictive models from multi-modal data, including data available at different levels of granularity or abstraction, across a broad range of applications, including predictive modeling of health outcomes and disease phenotypes from behavioral, clinical, biomedical, environmental, and socio-demographic data.

Topics in Grammar Inference and Computational Learning Theory

Grammatical Inference, variously referred to as automata induction, grammar induction, and automatic language acquisition, refers to the process of learning of grammars and languages from data. Machine learning of grammars finds a variety of applications in syntactic pattern recognition, adaptive intelligent agents, diagnosis, computational biology, systems modeling, prediction, natural language acquisition, data mining and knowledge discovery.

Our work on learning Regular Grammars demonstrates the feasibility of learning regular languages from examples under additional assumptions concerning the distribution from which the examples are drawn, thereby addressing the problem ("Are DFA PAC-identifiable if examples are drawn from the uniform distribution, or some other known simple distribution?") posed by Pitt in his seminal paper (against the background of strong negative results regarding the feasibility of learning regular grammars within the standard PAC learning framework):

(a) The class of simple DFA (i.e., DFA whose canonical representations have logarithmic Kolmogorov complexity) is efficiently PAC learnable under the Solomonoff Levin universal distribution (Parekh and Honavar, 1999).

(a) If the examples are sampled at random according to the universal distribution by a teacher that is knowledgeable about the target concept, the entire class of DFA is efficiently PAC learnable under the universal distribution, that is, DFA are efficiently learnable under the PACS Model (Parekh and Honavar, 1999; Parekh and Honavar, 2001).

(b) Any concept that is learnable under Gold's model for learning from characteristic samples, Goldman and Mathias' polynomial teachability model, and the model for learning from example based queries is also learnable under the PACS model (Parekh and Honavar, 2000; 2001).

Related work has led to the development of polynomial algorithms for learning regular languages from examples and membership queries (Nichitiu et al., 2000).

Our work on learning of grammars used to model natural languages (in particular, dependency grammars, stochastic context free grammars) has led to:

(a) The development of a novel regularization scheme, namely, *unambiguity regularization* that favors grammars that yield unambiguous parses, which includes as special cases and improves

upon, standard expectation maximization (EM), Viterbi EM, and Softmax EM algorithms for unsupervised learning of grammars (Tu and Honavar, 2012).

(b) Demonstration (and explanation) of the benefits of *curricula* (e.g., a means of ordering training samples presented to the learner in an inductive learning setting) using an incremental construction hypothesis which asserts (loosely speaking) that when the target of learning is a structure e.g., a grammar that can be decomposed into a set of sub-structures e.g., grammar rules, an ideal curriculum is one that gradually emphasizes data samples that help the learner to successively discover new substructures (Tu and Honavar, 2011).

(c) An iterative bi-clustering approach to learning probabilistic context free grammars (Tu et al., 2008, 2011).

Work in progress is aimed at extending the theoretical foundations and algorithms for grammar inference to settings that call for learning from multimodal data (e.g., combination of words and pictures). Some results to date include an algorithm for learning a multi-modal hierarchical Dirichlet process model for annotating images from partially labeled data (Yakhnenko and Honavar, 2009).

Explaining Machine Learned Predictive Models and their predictions

Our ability to acquire and annotate increasingly large amounts of data together with rapid advances in machine learning have made predictive models trained using machine learning ubiquitous in virtually all areas of human endeavor. In high-stakes applications such as healthcare, finance, criminal justice, scientific discovery, education, and others, the resulting predictive models are complex, and in many cases, black-boxes. Consider for example, a medical decision-making scenario where a predictive model, e.g., a deep neural network, trained on a large database of labeled data, is to assist physicians in diagnosing patients. In this setting, it is important that the clinical decision support system be able to explain the output of the deep neural network to the physician, who may not have a deep understanding of machine learning. For example, the physician might want to understand the subset of patient characteristics that contribute to the diagnosis; or the reason as to why diagnoses were different for two different patients, etc. In high stakes applications of machine learning, the ability to explain the machine learned model is a prerequisite for establishing trust in the model's predictions. Satisfactory explanations have to provide answers to questions such as: "What features of the input are responsible for the predictions?"; "Why are the model's outputs different for two individuals?" (e.g., Why did John's loan application get approved when Sarah's was not?). Hence, satisfactory explanations have to be fundamentally causal in nature. This project aims to develop a theoretically sound, yet practical approach to causal attribution, that is, apportioning the responsibility for a black-box predictive model's outputs among the model's inputs.

We reduce the model interpretation question "Why did the predictive model generate the output Y for input X?" to the following equivalent question: "How are the features of the model input X causally related to the model output Y?" In other words, we reduce the task of interpreting a black-box predictive model to the task of estimating, from observations of the inputs and the corresponding outputs of the model, the causal effect of each input variable or feature on the output variable. Our methods do not require knowledge of the internal structure or parameters of the black-box model, or of the objective function or the algorithm used to train the model. Hence, the resulting methods can be applied, in principle, to any black-box predictive model, so long as it is possible probe the model and observe the model's response to any supplied input data sample.

Federated Ontologies Knowledge Representation and Inference (funded in part by a grant from the National Science Foundation)

The success of the world wide web can be attributed to the *network effect*: The absence of central control on content and organization of the web allows thousands of independent actors to contribute resources (web pages) that are interlinked to constitute the web. Recent efforts to extend the web into a *semantic web* are aimed at enriching the web with machine interpretable content and interoperable resources and services. Realizing the full potential of the semantic web requires the large-scale adoption and use of

ontology based approaches to sharing of information and resources. In such a setting, instead of a single, centralized ontology, it is much more natural to have a federation of ontologies that cover different, perhaps partially overlapping, domains (e.g., biology, medicine, pharmacology). Such ontologies represent the *local* knowledge of the ontology designers, that is, knowledge that is applicable within a specific *context*. Hence, there is an urgent need for theoretically sound yet practical approaches that support user, context, or application-specific adaptation and reuse of knowledge from multiple autonomously developed ontologies in specific applications. Ontologies on the semantic web need to satisfy apparently conflicting objectives: Selective *sharing* or *r*euse of knowledge across autonomously developed ontologies on the one hand and accommodation of the *local points of view* or *contextuality* of knowledge on the other. Our research on modular ontologies has led to:

(a) The development of modular variants of description logics that provide mechanisms for *semantic* importing of names (including concept, role and nominal names) across ontology modules [Bao et al., 2006a; 2006b; 2007; 2009].

(b) Decidability and undecidability results for modular ontology languages establishing that (a) variants of distrtibuted description logics (DDL) that allow negated roles or cardinality restrictions in bridge rules or inverse bridge rules that connect *ALC* ontologies are undecidable;

(b) a variant of P-DL *ALCHIO*(←)*P* that supports role mappings between ontology modules in *ALCHIO*(¬) (an extension of *ALC* that allows general role inclusions, inverse roles, and negated roles) is decidable [Bao et al., 2008].

(c) The development of distributed tableau-based reasoning algorithms for P-DL [Bao, Caragea, and Honavar, 2006c; 2007b]

(d) Tools for collaborative development of package-based partial order ontologies for applications in biomedical informatics and comparative genomics [Hughes et al., 2008].

(e) Development of a sound and complete PSPACE tableau 27pprox.27 for the Description logic ALCK, i.e., Description Logic ALC augmented with Epistemic operators to support modeling of, and inference using, knowledge of multiple agents [Tao et al., 2012].

A long-term goal of our research on modular ontologies is to transform distributed data and knowledge base applications in the same way that the World-Wide Web has transformed the construction, sharing and use of hyperlinked documents and Wiki has transformed encyclopedia construction.

Secrecy-Preserving Inference and Query Answering (Funded in part by a grant from the National Science Foundation)

Productive interaction and collaboration among business partners, different governmental agencies (e.g., intelligence, law enforcement, public policy), or independent nations acting on matters of global concern (e.g., counter-terrorism, international finance) requires the need to share information to be balanced against the need to protect sensitive or confidential information from unintended disclosure. Our work, aimed at addressing this problem focuses on the theoretical foundations of, and algorithms and software for secrecy-preserving query answering, that is answering queries using secrets, whenever it is feasible to do so without revealing secrets. Results to date include:

(a) A framework for secrecy-preserving query answering in the simple case of hierarchical knowledge bases under open world semantics (Bao et al., 2007).

(b) A generalization of the secrecy-preserving query answering framework to a setting with multiple querying agents (each with a set of secrets that it is not permitted to know) that can pose queries against a knowledge base and selectively share answers received with one or more other querying agents. The framework exploits the indistinguishability of an answer that is not shared (because it is a secret) from an answer that is not entailed by the knowledge base. We showed how to use this approach in Propositional Horn and the Description Logic AL knowledge bases (Tao et al., 2015).

Work in progress is aimed at:

- The development of secrecy-preserving query answering systems for a knowledge bases of practical interest in networked information systems including hierarchical, propositional, RDF description logic (DL), and probabilistic knowledge bases.
- Examination of secrecy-preserving reasoning algorithms to settings with multiple querying agents, under various restrictions on communication among agents.

Representing and Reasoning About Qualitative Preferences (Funded in part by a grant from the National Science Foundation)

The ability to represent and reason about preferences over a set of alternatives (decisions, arguments, products, policies, etc.) is central to rational decision-making. In many real-world applications, preferences are best expressed in qualitative (as opposed to quantitative) terms. Alternatives are described in terms of a set of relevant attributes. Languages for representing preferences range from those that can express only unconditional preferences to those that can express conditional preferences, relative importance of preferences, etc. Our recent work (Santhanam et al., 2016) has led to:

(a) The first practical solution to the problem of determining whether one outcome dominates another with respect to a given set of qualitative preferences over the attributes of the outcomes (Santhanam et al., 2010). The proposed solution reduces the problem of dominance testing, to reachability analysis in a graph of outcomes. We provide an encoding of TCP-nets in the form of a Kripke structure for Computational Tree Logic (CTL). We show how to test dominance using a model checker for CTL.
(b) Specification of dominance relation that allows comparison of collections of objects in terms of preferences over attributes of the objects that make up the collection the set of most preferred collections, and algorithms for dominance testing that are guaranteed to return only (sound), all (complete), or at least one (weakly complete) of the most preferred collections (Santhanam et al., 2011).
(c) Applications of preference reasoning algorithms in service composition and substitution based on qualitative preferences over non-functional attributes of the services (Santhanam et al., 2008, 2009), and in minimizing credential disclosure based on qualitative preferences over sensitivity attributes of the credentials (Oster et al., 2013).

Work in progress is focused on:
- Extensions of the framework to handle reasoning with the preferences of multiple stakeholders;
- Applications of preference reasoning in product design, cyber-defense, and healthcare.

Algorithms and Software for Interactive Discovery and Composition of Web Services (Funded in part by a grant from the National Science Foundation).
Practical applications in a variety of domains, e.g., e-commerce, e-science, etc. increasingly rely on complex services that are composed from physically distributed and autonomously developed component services. Our recent work has focused on interactive methods for service composition with provable guarantees with respect to user-specified functional and non-functional requirements. Some results of this research include:
(a) Algorithms for interactive specification-driven functional assembly of composite services from a repository of available component services (Pathak et al., 2007a; 2007b; 2008)
(b) Algorithms for efficient identification of feasible replacement of one or more component services of a composite service while maintaining its functionality (Pathak et al., 2007)
(c) Algorithms for incorporating user preferences over non-functional attributes of a service (e.g., cost, security, reliability) in assembling a most preferred composition that achieves the user-specified functionality (Santhanam et al., 2008).

Macromolecular Sequence-Structure-Function-Interaction- Expression Relationships (in collaboration with Drena Dobbs and Robert Jernigan funded in part by a National Institutes of Health Grant 5R21GM066387)

Protein-protein, protein-DNA, and protein-RNA interactions play a pivotal role in virtually all biological processes. Reliably pinpointing which specific amino acid residues form the interface(s) between a protein and its binding partner(s) is critical for understanding the structural and physicochemical determinants of such interactions and has wide applications in modeling and validating interactions predicted by high-throughput methods, in engineering proteins, and in prioritizing drug targets. Because experimental determination of protein-protein and protein nucleic acid complexes is expensive, laborious, and error-prone, there is a need for development of reliable computational methods for identifying protein-protein interface residues. protein and its binding partner(s) is critical for understanding the structural and physicochemical determinants of such interactions and has wide applications in modeling and validating interactions predicted by high-throughput methods, in engineering proteins, and in prioritizing drug targets. Because experimental determination of protein-protein and protein nucleic acid complexes is expensive, laborious, and error-prone, there is a need for development of reliable computational methods for identifying protein-protein interface residues.

Against this background, our work aims to develop and systematically evaluate computational methods for computational characterization and prediction of protein-protein, protein-DNA, and protein-RNA interactions, interfaces, and complexes as well as residues and other functionally important sites (e.g., B-cell and T-cell epitopes, glycosylation and phosphorylation sites). Some of the results to date include:

(a) Comprehensive Database of Protein-protein Interfaces (Jordan et al., 2012) and of Protein-RNA Interfaces (Lewis et al., 2010).

(b) Development of a state-of-the-art approach to predicting protein-RNA interface residues in protein sequences (Walia et al., 2012, Terribilini et al., 2006, 2007; El-Manzalawy et al., 2016)

(c) Development of machine learning approaches and online servers for the prediction of protein- protein interface residues from amino acid sequence and when available, structural information (Yan et al., 2004; Towfic et al., 2011) including a state-of-the-art structure-based interface predictor (Jordan et al., 2012).

(d) Development of sequence homology based methods and online servers for protein interface prediction (Xue et al., 2011), including non partner-specific methods for predicting obligate interfaces and interfaces of disordered proteins and partner-specific methods for predicting transient interfaces.

(e) Development of sequence-based machine learning methods for predicting the approximate number of putative interaction partners of a protein (Andorf et al., 2013).

(f) Development of a novel approach and online server for scoring docked protein-protein complex conformations using predicted partner-specific protein-protein interfaces (Xue et al., 2011; 2013; 2016).

(g) Demonstration of the pitfalls of commonly used windows-based cross-validation for sequence- based classification tasks (e.g., phosphorylation site prediction, DNA-binding site prediction) (Caragea et al., 2009).

(h) Application of classifiers trained using machine learning to discover a large set of incorrect Gene Ontology annotations an experimentally well-studied family of proteins – mouse kinases (Andorf et al., 2007).

(i) Development of machine learning approaches and online servers for prediction of protein-DNA interface residues from amino acid sequence, and when available, structural information (Yan et al., 2006).

(j) Structural characterization of protein-protein and protein-RNA interfaces (Towfic et al., 2011).

(k) Development of machine learning methods and online servers for identification of posttranslational modification sites e.g., phosphorylation sites, glycosylation sites in amino-acid sequences (Caragea et al., 2007).

(l) Development of machine learning methods and online servers for predicting linear and B-cell epitopes from amino acid sequences (El-Manzalawy et al., 2008) including methods for predicting variable length and conformational B-cell epitopes.

(m) Demonstrations of the pitfalls of commonly used benchmark datasets for evaluating the performance of machine learning approaches to MHC-II binding site prediction (El-Manzalawy, 2008).

(n) Prediction of the designability of binary (H-P) protein sequences (Peto et al., 2008).

(o) Prediction of protein and RNA binding sites in recalcitrant (with regard to attempts at structure determination) proteins e.g., HIV-1 and EIAV and experimental confirmation of the predictions (with Lee et al., 2008).

The online web servers can be found at http://ailab.ist.psu.edu/software.html.

Comparative Analysis of Biomolecular Networks (funded in part by a grant from the USDA and in part by a National Science Foundation IGERT fellowship)

Network models play an increasingly important role in the interpretation of complex interactions among genes, proteins, regulatory RNAs, small ligands and other signaling agents. In particular, comparative analysis of network models of biomolecular interactions across different species or tissues has emerged as an important tool for identifying conserved modules, predicting functions of specific genes or proteins and studying the evolution of biological processes, among other applications. Hence, there is need for scalable modular, and extensible algorithms software for construction, querying, and comparative analysis of diverse types of biomolecular networks. Our work has led to:

(a) Development of a suite of modular graph kernel based scalable and customizable algorithms and their open source implementations for aligning protein-protein interaction networks and gene co-expression networks (Towfic et al., 2009). Graph kernels allow efficient computation of global alignment of networks by decomposing the task into a set of local graph kernel computations. The resulting software is available at http://ailab1.ist.psu.edu/BinaWebApp/.

(b) Application of comparative protein-protein interaction network analyses to reliably distinguish orthologs from paralogs (Towfic et al., 2010) and of comparative gene co-expression network analyses to identify B-cell ligand processing pathways (Towfic et al., 2012).

(c) Characterization of gene expression changes during the onset of photosynthesis (Lonosky et al., 2004), differentiation of retinal stem cells into rod photoreceptors (Hecker et al., 2010).

(d) Characterization differences in the proteome of murine retinal and brain derived progenitor cells (Dunn-Thomas et al., 2008).

(e) Development of databases and software tools for capture, analysis, annotation, and integration of gene expression data with other types of 'omics' data (Couture et al., 2009).

(f) Development of BioNetwork Bench, an open source, user-friendly suite of database and software tools for constructing, querying, and analyzing gene and protein network models (Kohutyuk et al., 2012; Hecker et al., 2008).

(g) Development and analysis of a machine learning algorithms for inference of temporal Boolean network models from multivariate time series data, with applications to inference of genetic networks from gene expression data (Silvescu and Honavar, 2001).

Work in progress is aimed at the:

- Further development of BiNA, our modular, extensible, and scalable suite of graph kernel based algorithms to enable alignment of richer networks including (a) Undirected graphs that contain multiple types of links (e.g., interaction, co-localization, etc. in the case of protein-protein interaction

networks), or multiple types of nodes (e.g., in the case of macromolecular interaction networks that simultaneously model the interactions among proteins, RNA, DNA, etc.), or both; (b) Directed graphs with one or more types of links (up or down regulation of one gene by another in the case of transcriptional networks), or one or more types of nodes, or both as in the case of richly annotated signaling networks and metabolic networks; (c) The weighted counterparts of undirected (e.g., gene expression correlation networks) as well as directed graphs and (d) Undirected or directed multi-graphs with multiple links between nodes as well as variants that accommodate sets of labels on nodes (e.g., Gene ontology functional annotation, subcellular localization, etc.) and links, as well as their weighted counterparts.

- Systematic evaluation of the graph kernel based network alignment algorithms (including comparisons them with existing algorithms) on several representative applications including: identifying differences in patterns of biomolecular interactions across different species or tissues; identifying conserved modules or subnetworks; predicting functions of specific genes or proteins (and identification of functional orthologs); compensating for limited experimental data concerning biomolecular interactions within one species through transfer of information from another species on several benchmark datasets.

Deciphering Brain Structure, Activity, and Function Relationships

The advent of new ways to determine brain structure and measure brain activity offer unprecedented opportunities to understand the complex relationship between how brains perform cognition and orchestrate behavior in naturalistic settings. Against this background, our work is aimed at developing and applying novel computational methods for:

- Understanding functional brain connectivity from brain activity (e.g., fMRI) and other data;
- Comparing brain networks across individuals, tasks, or conditions
- Characterizing and modeling longitudinal changes in brain networks due to development, learning, aging, and disease.

Biologically Inspired Algorithms for Knowledge Representation, Memory, Language Processing and Learning (Funded in part by a grant from the National Science Foundation)

Artificial neural networks, because of their potential for massive parallelism and fault and noise tolerance, offer an attractive approach to the design of associative memories, language processors, and trainable pattern classifiers. Constructive learning algorithms, which build arbitrarily complex decision boundaries needed for pattern classification (and in some ways, foreshadowed the recent development of support vector machines) were motivated by: the need to overcome the limitations of learning through parameter modification within an a priori fixed network topology; and to avoid the guesswork involved in deciding suitable network architectures for different pattern classification problems by dynamically growing the network to match the complexity of the underlying pattern classification task. Evolutionary algorithms offer a powerful means of exploring large search spaces for solutions that optimize multiple objectives e.g., feature subsets that maximize the predictive performance and minimize the complexity of the classifiers that use them. Against this background, we explored several closely related topics in biologically inspired (neural, evolutionary) algorithms and architectures for knowledge representation, language processing e.g., parsing, and learning. This work has led to:

(a) Generalization (with convergence guarantees) of a large family of constructive neural network learning algorithms designed for 2-class binary pattern classification problems to handle classification problems involving real-valued patterns and an arbitrary number of classes (Parekh et al., 2000).

(b) Development of a simple, inter-pattern distance based provably convergent, polynomial time constructive neural network algorithm which compares very favorably with computationally far more expensive algorithms in terms of generalization accuracy (Yang et al., 1999).

(c) Development of algorithms for construction of robust, noise-tolerant neural memories for pattern

storage and associative, content-based retrieval (Chen et al., 1995) and query processing(1996).

(d) Development of algorithms for construction of highly parallel neural architectures for syntax analysis (parsing of regular, context-free, and context-sensitive languages) (1999).

(e) Development of a biologically inspired neural architecture and an extended Kalman filter algorithm for place learning and localization in a-priori unknown environments which successfully accounts for a large body of behavioral and neurobiological data from animal experiments and offers several testable predictions (Balakrishnan et al., 1998, 2000).

(f) Development of evolutionary algorithms for feature subset selection for classification problems (Yang et al., 1998) and sensors and controllers for adaptive robots (Balakrishnan and Honavar, 1996; 1998; 2001).

(g) Development of incremental neural network learning algorithms with applications in sensing and nondestructive evaluation (Polikar et al., 2001a, 2001b, 2004).

(h) Development of constructive neural network algorithms that take advantage of prior knowledge in the form of classification rules (Parekh et al., 1999).

(i) Hybrid neural-symbolic architectures for information processing (Honavar and Uhr, 1994;1995).

Critical Infrastructure Monitoring and Protection (Funded in part by grants from the National Science Foundation and the US Department of Defense)

Formal methods offer a rigorous foundation for representing and reasoning about critical infrastructure for information and communication, transportation, and health. Data-driven approaches complement formal methods in domains where the available formal specifications are incomplete. Our work has used both approaches to solve problems that arise in monitoring and protecting critical infrastructure. This work has led to:

(a) Development of model checking approaches to containing infection propagation across networks (Santhanam et al., 2011).

(b) Development of tools for formal specification of intrusions software fault trees and of intrusion detection systems using colored Petri nets s for automated generation of multi-agent systems for coordinated intrusion detection in computer and communication networks (Helmer et al., 2002; 2006).

(c) Development of multi-agent system for detection of coordinated or concerted attacks on distributed computing systems in particular by monitoring different processes, resources, users, events, and extract and integrate relevant information from disparate sources over multiple space and time scales (Helmer et al., 2003Wang et al., 2006).

(d) Development and application of machine learning approaches for learning predictive rules for anomaly and misuse detection (Kang et al., 2005, 2006).

(e) Development of an electronic nose for detection and identification of odorants using machine learning (Polikar et al., 2001).

(f) Development and applications of machine learning methods for non-destructive inspection of nuclear power plant pipes using ultrasound (Polikar et al., 2006).

(g) Development of a service-oriented distributed software infrastructure for monitoring distributed power systems (Pathak et al., 2006; 2007; McCalley et al., 2007; Pham et al., 2009).

(h) Design, analysis, and evaluation of a utility-theoretic approach to routing in communication networks that supports a flexible tradeoff between delay for a specific message and the overall network load (and hence expected delay for all routed messages) using a knowledge representation scheme that enables each node in a communication network to maintain and update a small constant-size knowledge base (independent of the network size) (Mikler et al., 1996; 1997; 1998; 2001).

PUBLICATIONS AND PRESENTATIONS

Books Authored and Edited

1. Santhanam, G., Basu, S., and **Honavar, V.** (2016). Representing and Reasoning About Qualitative Preferences. Synthesis Lectures in Artificial Intelligence and Machine Learning. Morgan and Claypool.
2. Patel, M., **Honavar, V**. & Balakrishnan, K. (Ed.) (2001). **Advances in Evolutionary Synthesis of Intelligent Agents***.* Cambridge, MA: MIT Press.
3. Honavar, V. & Uhr, L. (1994) (Ed). **Artificial Intelligence and Neural Networks***:* Steps Toward Principled Integration*.* New York, NY: Academic Press.

Influential White Papers

1. **Honavar, V**., Hill, M., and Yelick, K. (2016). Accelerating Science: A Computing Research Agenda. A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. arXiv preprint arXiv:1604.02006.
2. Hager G., Bryant R., Horvitz E., Matarić M., & **Honavar V.**, (2017). Advances in Artificial Intelligence Require Progress Across all of Computer Science: http://cra.org/ccc/wp-content/uploads/sites/2/2015/01/CCC-AI-Systems-2017-FINAL.pdf
3. Barocas, S., Bradley, E., **Honavar, V.** and Provost, F. (2017). Big Data, Data Science, and Civil Rights A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. arXiv preprint arxiv:1706.03102.
4. **Honavar, V.**, Yelick, K., Nahrstedt, K., Rushmeier, H., Rexford, J., Hill, Mark., Bradley, E., and Mynatt, E. (2017). Advanced Cyberinfrastructure for Science, Engineering, and Public Policy. A white paper prepared for the Computing Community Consortium committee of the Computing Research Association. arXiv preprint arXiv:1707.00599.

Refereed Journal and Conference Papers

Note: Acceptance of papers for publication in the proceedings of top Computer Science conferences is typically based on rigorous peer review. Acceptance rates are typically under 30% and conference proceedings are published and distributed by major commercial publishers (e.g., Springer-Verlag) or professional societies (e.g., Association for Computing Machinery). Dr. Honavar's work (which includes one of the 10 most cited papers in IEEE Intelligent Systems), according to Google Scholar Data (as of August 2020), has been cited more than 15,500 times (or over 500 citations on average per year during 1990-2020) and his *h*-index[2] is 58, *g*-index is 86, and *i*10-index is 234.

5. Liang, J., Wu, Y., Yu, D., and **Honavar, V.** (2021). Longitudinal Deep Kernel Gaussian Process Regression In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. In press.
6. Hsieh, T-Y., Sun, Y., Wang, S. and **Honavar, V.** (2021). Explainable Multivariate Time Series Classification: A Deep Neural Network Which Learns to Attend To Important Variables As Well As Informative Time Intervals. In: Proceedings of the 14th International Conference on Web Search and Data Mining. In press.
7. Liang, J.(*), Guo, W.(*), Luo, T., **Honavar, V.**, Wang, G., Xing, X. (2021) "FARE: Enabling Fine-grained Attack Categorization under Low-quality Labeled Data", Proceedings of The Network and Distributed System Security Symposium (NDSS), In press.
8. Hsieh, T-Y., Sun, Y., Wang, S., and **Honavar, V.** (2021). Functional Autoencoders for Functional

Data Representation Learning. In: Proceedings of the SIAM Conference on Data Mining. In press.

9. Hsieh, T-Y., Sun, Y., Tang, X., Wang, S., and **Honavar, V**. (2021). SRVAR: Joint Discrete Hidden State Discovery and StructureLearning from Time Series Data. In: Proceedings of the Web Conference. In press.

10. Seto, C., Khademi, A., Graif, C., and **Honavar, V.** (2021) "Commuting Network Spillovers and COVID-19 Deaths Across US Counties." In: The Annual Meeting of the Population Society of America. In press.

11. Le. T. and **Honavar, V.** (2020). Dynamical Gaussian Process Latent Variable Model for Representation Learning from Longitudinal Data Proceedings of the 2020 ACM-IMS Conference on the Foundations of Data Science pp. 183-188. https://doi.org/10.1145/3412815.3416894

12. Liang, J., Xu, D., Sun, Y., and **Honavar, V.** (2020). LMLFM: Longitudinal Multi-Level Factorization Machines. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence.

13. Hou, Y., Wu, C., Yang, D., Ye, T., **Honavar, V.G.**, van Duin, A.C., Wang, K. and Priya, S. (2020). Two-dimensional hybrid organic–inorganic perovskites as emergent ferroelectric materials. Journal of Applied Physics, *128*(6), p.060906.

14. Renaud, N., Jung, Y., **Honavar, V**., Geng, C., Bonvin, A.M. and Xue, L.C. (2020). iScore: An MPI supported software for ranking protein–protein docking models based on a random walk graph kernel and support vector machines. *SoftwareX*, *11*, p.100462.

15. Le, T. and Honavar, V. (2020) Dynamical Gaussian Process Latent Variable Model for Representation Learning from Longitudinal Data. In: Proceedings of the ACM/IMS Conference on Foundations of Data Science.

16. Lee, S. and **Honavar, V**., (2020), Towards robust relational causal discovery. In Uncertainty in Artificial Intelligence (pp. 345-355). PMLR.

17. Parashar, M., **Honavar, V**., Simonet, A., Rodero, I., Ghahramani, F., Agnew, G., and Jantz, R. (2020). The Virtual Data Collaboratory: A Regional Cyberinfrastructure for Collaborative Data-Driven Research. Computing in Science & Engineering, 22(3), pp.79-92.

18. Sun, Y., Wang, S., Tang, X., Hsieh, T-Y., and **Honavar, V.** (2020). Non-target-specific Node Injection Attacks on Graph Neural Networks: A Hierarchical Reinforcement Learning Approach. Proceedings of The Web Conference 2020 (WWW '20) https://doi.org/10.1145/3366423.3380149

19. Geng, C., Jung, Y., Renaud, N., **Honavar, V.**, Bonvin, A., Xue, L. (2020). iScore: A novel graph kernel-based function for scoring protein-protein docking models, Bioinformatics. https://doi.org/10.1093/bioinformatics/btz496.

20. Kandasamy, S., Bhattacharyya, A., and **Honavar, V.** (2019). Minimum Intervention Cover of a Causal Graph. In: Proceedings of the 33rd AAAI Conference on Artificial Intelligence.

21. Khademi, A., Lee, S., Foley, D., and **Honavar, V.** (2019). Fairness in Algorithmic Decision Making: A Preliminary Excursion Through the Lens of Causality. In: Proceedings of the Web Conference.

22. Zhou, Y., Sun, Y., and **Honavar, V.** (2019). Improving Image Captioning Using Knowledge Graphs. In: Proceedings of the Winter Conference on Applications of Computer Vision. IEEE Press.

23. Jung, Y., EL-Manzalawy, Y., Dobbs, D., & **Honavar, V.** (2018). Partner-specific Prediction of RNA-binding Residues in Proteins: A Critical Assessment. Proteins: Structure, Function, and Bioinformatics. http://doi.org/10.1002/prot.25639

24. Liang, J., Hu, J., Dong, S., and **Honavar, V.** (2018). Top-N-Rank: A Truncated List-wise Ranking Approach for Large-scale Top-N Recommendation. In: Proceedings of the IEEE International

Conference on Big Data.

25. Hsieh, T-Y., El-Manzalawy, Y., Sun, Y., and **Honavar, V** (2018). Compositional Stochastic Average Gradient for Machine Learning and Related Applications. In: Proceedings of the 19th International Conference on Intelligent Data Engineering and Automated Learning. **Best paper award.**

26. Abbas M, Le T, Bensmail H, **Honavar V**, El-Manzalawy Y (2018). Microbiomarkers Discovery in Inflammatory Bowel Diseases using Network-Based Feature Selection. Proceedings of the 9th ACM Conference on Bioinformatics, Computational Biology and Health Informatics.

27. El-Manzalawy, Y., Hsieh, T.Y., Shivakumar, M., Kim, D. and **Honavar, V**. (2018). Min-Redundancy and Max-Relevance Multi-view Feature Selection for Predicting Ovarian Cancer Survival using Multi-omics Data. BMC Medical Genomics.

28. Hu, J, Liang, J, Kuang, Y, and **Honavar, V.** (2018). A user similarity-based Top-N recommendation approach for mobile in-application advertising Expert Systems with Applications, vol. 111, pp. 51-60. DOI: 10.1016/j.eswa.2018.02.012

29. Gur, S., and **Honavar, V**. (2018). PATENet: Pairwise Alignment of Time Evolving Networks.. in: Proceedings of the 14th International Conference on Machine Learning and Data Mining in Pattern Recognition. Lecture Notes in Computer Science, vol. 10934 LNAI, Springer Verlag, pp. 85-98

30. Khademi, A., El-Manzalawy, Y., Buxton, O., and **Honavar, V.** (2018). Toward Personalized Sleep/Wake Prediction from Actigraphy. IEEE International Conference on Biomedical and Health Informatics. pp. 414-417.

31. Sun, Y., Bui, N., Hsieh, T-Y., and **Honavar, V**. (2018). Multi-View Network Embedding Via Graph Factorization Clustering and Co-Regularized Multi-View Agreement. In: Proceedings of the Eighth IEEE ICDM Workshop on Data Mining in Networks, IEEE.

32. Vairo, G. L., Conroy, D. E., & **Honavar, V.** (2018). Acceptability of Personal Sensing to Develop a Digital Biomarker of Stress-Related Injury Risk in Athletes. Medicine & Science in Sports & Exercise, 50, 605. http://http://dx.doi.org/10.1249/01.mss.0000537080.70348.c2.

33. Lee, S. and **Honavar, V**. (2017). Self-Discrepancy Conditional Independence Test. In: Conference on Uncertainty in Artificial Intelligence.

34. Lee, S. and **Honavar, V.** (2017). A Kernel Independence Test for Relational Data. In: Conference on Uncertainty in Artificial Intelligence.

35. El-Manzalawy, Y., Buxton, O., and **Honavar, V**. (2017). Sleep/wake state prediction and sleep parameter estimation using unsupervised classification via clustering. In: IEEE Conference on Bioinformatics and Biomedicine.

36. El-Manzalawy, Y., Hsieh, T-Y., Shivkumar, M., Kim, D., and **Honavar, V**. (2017). Min-Redundancy and Max-Relevance Multi-view Feature Selection for Predicting Ovarian Cancer Survival using Multi-omics Data. In: The 7th Annual Translational Bioinformatics Conference.

37. El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2017). In silico prediction of linear B-cell epitopes on proteins. In: Y. Zhou, E. Faraggi, A. Kloczkowski and Y. Yang (Eds.), Prediction of Protein Secondary Structure, Methods in Molecular Biology, vol. 1484, DOI:10.1007/978-1- 4939-6406-2_17.

38. Walia, R., El-Manzalawy, Y., Dobbs, D., and **Honavar, V**. (2017). Sequence-based Prediction of RNA-binding Residues in Proteins. Methods in Molecular Biology, vol. 1484, DOI:10.1007/978- 1-4939-6406-2_15.

39. Xue, L., Rodrigues, J.P.L.M., Dobbs, D., **Honavar, V**., and Bonvin, A. (2017). Template-Based Protein-Protein Docking Improved Using Pairwise Interfacial Residue Restraints. Briefings in Bioinformatics doi: 10.1093/bib/bbw027

40. Bui, N., Le, T. and **Honavar, V**. (2016). Labeling Actors in Multi-view Social Networks by Integrating Information From Within and Across Multiple Views. In: Proceedings of the IEEE Conference on Big Data.

41. Bui, N., Yen, J., and **Honavar, V.** (2016). Temporal Causality Analysis of Sentiment Change in a Cancer Survivor Network. IEEE Transactions on Computational Social Systems. doi:10.1109/TCSS.2016.2591880.

42. Lee, S. and **Honavar, V.** (2016). A Characterization of Markov Equivalence Classes of Relational Causal Models under Path Semantics In: Proceedings of the Conference on Uncertainty in Artificial Intelligence (UAI 2016), pp. 387-396.

43. Lee, S. and **Honavar, V.** (2016). On learning causal models from relational data. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI-16). pp.3262- 3270.

44. El-Manzalawy, Y., Munoz, E., Lindner, S.E., and **Honavar, V.** (2016). PlasmoSEP: Predicting surface-exposed proteins on themalaria parasite using semisupervised self-training and expert-annotated data. Proteomics. doi: DOI 10.1002/pmic.201600249.

45. El-Manzalawy, Y., Abbas, M., Malluhi, Q., and **Honavar, V.** (2016). FastRNABindR: Fast and Accurate Prediction of Protein-RNA Interface Residues. PLoS One 11(7): e0158445. doi:10.1371/journal.pone.0158445

46. El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2016). In silico prediction of linear B-cell epitopes on proteins. In: Y. Zhou, E. Faraggi, A. Kloczkowski and Y. Yang (Eds.), Prediction of Protein Secondary Structure, Methods in Molecular Biology, vol. 1484, DOI:10.1007/978-1- 4939-6406-2_17.

47. Walia, R., El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2016). Sequence-based Prediction of RNA-binding Residues in Proteins. In: Y. Zhou, E. Faraggi, A. Kloczkowski and Y. Yang (Eds.), Prediction of Protein Secondary Structure, Methods in Molecular Biology, vol. 1484, DOI:10.1007/978-1-4939-6406-2_15.

48. Xue, L., Rodrigues, J.P.L.M., Dobbs, D., **Honavar, V.**, and Bonvin, A. (2016). Template-Based Protein-Protein Docking Improved Using Pairwise Interfacial Residue Restraints. Briefings in Bioinformatics doi: 10.1093/bib/bbw027

49. Bui, N., Yen, J. and **Honavar, V**. (2015). Temporal Causality of Social Support in an Online Community for Cancer Survivors In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP15). Springer-Verlag Lecture Notes in Computer Science, Vol. 9021, pp. 13-23.

50. Lee, S., and **Honavar, V.** (2015). Lifted Representation of Relational Causal Models Revisited: Implications for Reasoning and Structure Learning In: Workshop on Advances in Causal Inference, Conference on Uncertainty in Artificial Intelligence, 2015.

51. Lin, H., Bui, N., and **Honavar, V.** (2015). Learning Classifiers from Remote RDF Data Stores Augmented with RDFS Subclass Hierarchies. In: 2nd International Workshop on High Performance Big Graph Data Management, Analysis,and Mining (BigGraph 2015), The IEEE International Conference on Big Data.

52. Sawyer, A., Kang, Y., **Honavar, V**., Griffin, P., and Prabhu, V. (2015). Stimulating new and innovative perspectives on old and persistent problems: A commentary on Wohlgemuth, et al. "Attempters, adherers, and non-adherers: Latent profile analysis of CPAP use with correlates. Sleep Medicine, Vol. 16. pp. 311-312.

53. Xue, L., Dobbs, D., Bonvin, A., and **Honavar, V.** (2015). Computational Prediction of Protein Interfaces: A Review of Data Driven Methods. FEBS Letters. Vol. 589, No. 23, pp. 3516-3526.

54. Bui, N. and **Honavar, V.** (2014). Labeling Actors in Social Networks Using a Heterogeneous Graph Kernel. In: International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction (SBP14). pp. 27-34.

55. Walia, RR., Xue, LC., Wilkins, K., El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2014). RNABindRPlus: A Predictor that Combines Machine Learning and Sequence Homology-Based Methods to Improve the Reliability of Predicted RNA-Binding Residues in Proteins, PloS one 9 (5), e97725

56. Xue, L., Jordan, R., El-Manzalawy, Y., Dobbs, D., and **Honavar, V**. (2014) DockRank: Ranking Docked Conformations Using Partner-Specific Sequence Homology Based Protein Interface Prediction. Proteins: Structure, Function and Bioinformatics. Vol. 82, pp. 250-267. DOI: 10.1002/prot.24370.

57. **Honavar, V.** (2014). The Promise and Potential of Big Data: A Case for Discovery Informatics Review of Policy Research. In press.

58. Andorf, C., **Honavar, V**. and Sen, T. (2013). Predicting the Binding Patterns of Proteins: A Study Using Yeast Protein Interaction Networks. PLOS One 8(2): e56833, doi: 10.1371/journal.pone.0056833

59. Bareinboim, E., Lee, S., **Honavar, V.** & Pearl, J. (2013). Transportability from Multiple Environments with Limited Experiments. In: Advances in Neural Information Systems (NIPS) 2013. pp. 136-144.

60. Bui, N. and **Honavar, V.** (2013). On the utility of abstraction in labeling actors in social networks. In: Proceedings of the International Conference on Advances in Social Network Analysis and Mining (ASONAM 2013). pp. 692-698.

61. Lee, S. and **Honavar, V.** (2013). m-Transportability: Transportability of a Causal Effect from Multiple Environments. In: Proceedings of the 27th Conference on Artificial Intelligence (AAAI 2013). pp. 583-590.

62. Lee, S. and **Honavar, V**. (2013). Causal Transportability of Experiments on Controllable Subsets of Variables: z-Transportability. In: Proceedings of the 29th Conference on Uncertainty in Artificial Intelligence (UAI 2013).

63. Lin, H. and **Honavar, V.** (2013). Learning Classifiers from Chains of Multiple Interlinked RDF Data Stores. In: IEEE Big Data Congress. **Best Student Paper Award**. pp. 94-101.

64. Lin, H., Lee, S., Bui, N. and **Honavar, V**. (2013). Learning Classifiers from Distributional Data. In: IEEE Big Data Congress. pp. 302-309.

65. Kumar, S, Nilsen, W., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., Riley, W. T, Shar, A., Spring, B., Spruijt-Metz, D., Hedeker, D, **Honavar, V**., Kravitz, R.L., R. Lefebvre, C., Mohr, D.C., Murphy, S.A., Quinn, C., Shusterman, V., Swendeman, D. (2013). Exploring Innovative Methods to Evaluate the Efficacy and Safety of Mobile Health. American Journal of Preventive Medicine, 45(2): 228-236.

66. Oster, Z., Santhanam, G., Basu, S. and **Honavar, V.** (2013). Model Checking of Qualitative Sensitivity Preferences to Minimize Credential Disclosure. International Symposium on Formal Aspects of Component Software. Springer-Verlag LNCS Vol. 7684, pp. 205-223.

67. Letao Qi, Harris T. Lin, **Honavar, V** (2013): Clustering remote RDF data using SPARQL update queries. In: The 4th International Workshop on Graph Data Management: Techniques and Applications (GDM 2013), ICDE Workshops 2013: 236-242

68. Santhanam, G. Basu S., and **, Honavar, V**. (2013). Verifying Preferential Equivalence and Subsumption via Model Checking. Proceedings of the Conference on Algorithmic Decision Theory. Springer-Verlag LNCS Vol. 8176 pp. 324-335

69. Silvescu, A., and **Honavar, V**. (2013). Abstraction Super-Structuring Normal Forms: Towards a Theory of Structural Induction. In: Solomonoff Festschrift, Springer-Verlag LNAI Vol. 7070. Pp. 339-350.

70. Yasser El-Manzalawy, Drena Dobbs, **Vasant Honavar** (2012): Predicting protective bacterial antigens using random forest classifiers. In Proceedings of the ACM Conference on Bioinformatics and Computational Biology (BCB 2012), pp. 426-433

71. Jordan, R., El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2012). Predicting protein-protein interface residues using local surface structural similarity. BMC Bioinformatics 2012, 13:41 doi:10.1186/1471-2105-13-41. Highly Accessed.

72. Tao, J., Slutzki, G., and **Honavar, V.** (2012). Pspace Tableau Algorithms for Modalized ALC. Journal of Automated Reasoning. DOI: 10.1007/s10817-011-9232-3, 49(4): 551-582.

73. Towfic, F., Gupta, S., **Honavar, V**., and Subramaniam, S. (2012). B-Cell Ligand Processing Pathways Detected by Large-Scale Gene Expression Analysis. Genomics, Proteomics, and Bioinformatics. 10(3): 142-152 (2012).

74. Towfic, F., Kohutyuk, O., Greenlee, MHW., and **Honavar, V**. (2012). Bionetworkbench: Database and Software for Storage, Query, and Interactive Analysis of Gene and Protein Networks. Bioinformatics and Biology Insights 2012:6 235-246, doi: 10.4137/BBI.S9728

75. Tu, K. and **Honavar, V.** (2012). Unambiguity Regularization for Unsupervised Learning of Probabilistic Grammars. In: Proceedings of EMNLP-CoNLL 2012 : Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. Pp. 1324-1334

76. Walia, R., Caragea, C., Lewis, B., Towfic, F., Terribilini, M., El-Manzalawy, Y., Dobbs, D., **Honavar, V**. (2012). Protein-RNA Interface Residue Prediction Using Machine Learning: An Assessment of the State of the Art. BMC Bioinformatics 13:89 doi:10.1186/1471-2105-13-89. **Highly Accessed.**

77. El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2011). Predicting MHC-II binding affinity using multiple instance regression. IEEE/ACM Transactions on Computational Biology and Bioinformatics. DOI: 10.1109/TCBB2010.94

78. Lewis, B.A., Walia, R.R., Terribilini, M., Ferguson, J., Zheng, C., **Honavar, V.,** and Dobbs, D. (2011). PRIDB: A Protein-RNA Interface Database. Nucleic Acids Research. D277-282. DOI: 10.1093/nar/gkq1108.

79. Lin, H., Koul, N., and **Honavar, V**. (2011). Learning Relational Bayesian Classifiers from RDF data. In: Proceedings of the International Semantic Web Conference (ISWC 2011). Springer-Verlag Lecture Notes in Computer Science Vol. 7031 pp. 389-404.

80. Muppirala, U.K., **Honavar, V**., Dobbs, D. (2011). Predicting RNA-Protein Interactions Using Only Sequence Information. BMC Bioinformatics 2011, 12:489, doi:10.1186/1471-2105-12-489. **Highly Accessed.**

81. Santhanam, G., Basu, S., and **Honavar, V.** (2011). Representing and Reasoning with Qualitative Preferences for Compositional Systems. Journal of Artificial Intelligence Research Vol 42, pp. 211-274.

82. Santhanam, G., Suvorov, Y., Basu, S., and **Honavar, V.** (2011). Verifying Intervention Policies for Infection Propagation over Networks: A Model Checking Approach. In: Proceedings of the Twenty-Fifth Conference on Artificial Intelligence (AAAI-2011). Pp. 1408-1414.

83. Santhanam, G., Basu, S., and **Honavar, V.** (2011). Identifying Sustainable Designs Using Preferences Over Sustainability Attributes. In: AAAI Spring Symposium on Artificial Intelligence. Pp. 91-97.

84. Silvescu, A. and **Honavar, V.** (2011). Abstraction Super-structuring Normal Forms: Towards a Theory of Structural Induction. In: The Proceedings of the Solomonoff 85th Memorial Conference. Springer-Verlag Lecture Notes in Artificial Intelligence. In press.

85. Tu, K. and **Honavar, V.** (2011). On the Utility of Curricula in Unsupervised Learning of Grammars. In: Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011) pp. 1523-1528.

86. Tu, K., Ouyang, X., Han, D., Yu, Y., and **Honavar, V**. (2011). Exemplar-based Robust Coherent Biclustering. In: Proceedings of the SIAM Conference on Data Mining (SDM 2011). Pp. 884-895.

87. Xue, L., Dobbs, D., and **Honavar, V**. (2011). HomPPI: A Class of Sequence Homology Based Protein-Protein Interface Prediction Methods. BMC Bioinformatics 12:244 doi:10.1186/1471-2105-12-244

88. Xue, L., Jordan, R., El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2011). Sequence Based Partner-Specific Prediction of Protein- Protein Interfaces and its Application in Ranking Docked Models. In: ACM Conference on Bioinformatics and Computational Biology.

89. Yakhnenko, O., and **Honavar, V**. (2011). Multi-Instance Multi-Label Learning for Image Classification with Large Vocabularies. In: Proceedings of the British Machine Vision Conference. http://dx.doi.org/10.5244/C.25.59.

90. Barnhill, A.E., Hecker, L.A., Kohutyuk, O., Buss, J.E., **Honavar, V**. and Greenlee, H.W. (2010) Characterization of the Retinal Proteome During Rod Photoreceptor Genesis. BMC Research Notes 3:25.

91. Caragea, C., Silvescu, A., Caragea, D. and **Honavar, V.** (2010). Abstraction-Augmented Markov Models. In: Proceedings of the IEEE Conference on Data Mining (ICDM 2010). IEEE Press. Pp. 68-77.

92. Caragea, C. Silvescu, A., Caragea, D., and **Honavar, V.** (2010). Semi-supervised prediction of protein subcellular localization using abstraction augmented Markov models. BMC Bioinformatics. Doi: 10.1186/1471-2105-11-S8-S6.

93. Caragea, C., Silvescu, A., Caragea, D., and **Honavar, V**. (2010). Semi-Supervised Sequence Classification Using Abstraction Augmented Markov Models. In: Proceedings of the ACM Conference on Bioinformatics and Computational Biology. Pp. 257-264, doi: 10.1145/1854776.1854813. ACM Press.

94. El-Manzalawy, Y. and **Honavar, V**. (2010). Recent Advances in B-Cell Epitope Prediction Methods. Immunome Research Suppl. 2:S2. http://www.immunome- research.com/content/pdf/1745-7580-6-S2-S2.pdf.

95. Koul, N., Bui, N., and **Honavar, V**. (2010). Scalable, Updatable Predictive Models for Sequence Data. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2010).

96. Koul, N. and **Honavar, V.** (2010). Learning in the Presence of Ontology Mapping Errors. In: Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. Pp. 291-296. ACM Press.

97. Pandit, S., and **Honavar, V**. (2010). Ontology-Guided Extraction of Complex Nested Relationships from Text. IEEE Conference on Tools With Artificial Intelligence (ICTAI 2010). Pp. 173-178.

98. Sanghvi, B., Koul, N., and **Honavar, V.** (2010). Identifying and Eliminating Inconsistencies in Mappings across Hierarchical Ontologies. In: Springer-Verlag Lecture Notes in Computer Science Vol. 6427, pp. 999-1008. Berlin: Springer.

99. Santhanam, G., Basu, S., and **Honavar, V**. (2010). Dominance Testing Via Model Checking. In: Proceedings of the 24th AAAI Conference on Artificial Intelligence (AAAI-10). AAAI Press. Pp. 357-362.

100. Santhanam, G., Basu, S., and **Honavar, V**. (2010). Efficient Dominance Testing for Unconditional Preferences. In: Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010), pp 590-592.

101. Sun, H., Basu, S., **Honavar, V**., and Lutz, R. (2010). Automata-Based Verification of Security Requirements of Composite Web Services. In: Proceedings of the IEEE International Symposium on Software Reliability Engineering (ISSRE-2010). Pp. 348-357, IEEE Press.

102. Tao, J., Slutzki, G., and **Honavar, V.** (2010). Secrecy-preserving Query Answering for EL. In: Proceedings of the 4th International Conference on Web Reasoning and Rule Systems (RR 2010). Berlin: Springer. In press.

103. Towfic, F., VanderPlas, S., Oliver, C.A., Couture, O., Tuggle, C.K., Greenlee, M.H.W., and **Honavar, V.** (2010). Detection of gene orthology from gene co-expression and protein interaction networks. BMC Bioinformatics, BMC Bioinformatics 2010, 11 (Suppl 3): S7 doi:10.1186/1471-2105-11-S3-S7

104. Towfic, F., Caragea, C., Dobbs, D., and **Honavar, V**. (2010). Struct-NB: Predicting protein-RNA binding sites using structural features. International Journal of Data Mining and Bioinformatics, Vol. 4, No.1 pp. 21 – 43

105. Tuggle, C.K., Bearson, S.M.D, Huang, T.H., Couture, O., Wang, Y., Kuhar, D., Lunney, J.K., **Honavar, V.** (2010). Methods for transcriptomic analyses of the porcine host immune response: Application to Salmonella infection using microarrays. Veterinary Immunology and Immunopathology. Vol. 138. Pp. 282-291.

106. Bao, J., Voutsadakis G., Slutzki, G. **Honavar, V.** (2009). Package-Based Description Logics. In: Modular Ontologies: Concepts, Theories and Techniques for Knowledge Modularization. Lecture Notes in Computer Science Vol. 5445, pp. 349-371.

107. Bromberg, F., Margaritis, D., and **Honavar, V.** (2009). Efficient Markov Network Structure Discovery from Independence Tests. Journal of Artificial Intelligence Research. Vol. 35. Pp. 449-485.

108. Caragea, C., Sinapov, J., Dobbs, D., and **Honavar, V**. (2009). Mixture of experts models to exploit global sequence similarity on biomolecular sequence labeling. BMC Bioinformatics. Doi:10.1186/1471-2105-10-S4-S4.

109. Caragea, C., Caragea, D., and **Honavar, V**. (2009). Learning Link-Based Classifiers from Ontology-Extended Distributed Data. In: Proceedings of the 8th International Conference on Ontologies, Databases, and Applications of Semantics (ODBASE). Springer-Verlag Lecture Notes in Computer Science vol. 5871, pp. 1139-1146.

110. Caragea, C., Caragea, D., and **Honavar, V.** (2009). Learning Link-Based Classifiers from Ontology-Extended Textual Data. In: Proceedings of the IEEE Conference on Tools with Artificial Intelligence. IEEE Press. Pp. 354-361

111. Couture, O., Callenberg, K., Koul, N., Pandit, S., Younes, J., Hu, Z-L., Dekkers, J., Reecy, J., **Honavar, V**., and Tuggle, C. (2009). ANEXdb: An Integrated Animal Annotation and Microarray Expression Database. Mammalian Genome. Vol. 20. Pp. 768-777.

112. El-Manzalawy, Y. and **Honavar, V**. (2009). MICCLLR: Multiple-Instance Learning using Class Conditional Log Likelihood Ratio. In: Proceedings of the 12th International Conference on Discovery Science (DS 2009). Springer-Verlag Lecture Notes in Computer Science vol. 5808. Pp. 80-91.

113. Koul, N., and **Honavar, V**. (2009). Design and Implementation of a Query Planner for Data Integration. In: Proceedings of the IEEE Conference on Tools with Artificial Intelligence. IEEE Press. Pp. 214-218.

114. Pham, H., Santhanam, G., McCalley, J., and **Honavar, V**. (2009). BenSOA: a Flexible Service-Oriented Architecture for Power System Asset Management. In Proceedings of the North American Power Symposium (NAPS).

115. Santhanam, G.R., Basu, S., and **Honavar, V**. (2009). Web Service Substitution Based on Preferences Over Non-functional Attributes. In: Proceedings of the IEEE International Conference on Services Computing (SCC 2009). IEEE Press. Pp. 210-217.

116. Silvescu, A., Caragea, C., and **Honavar, V**. (2009). Combining Abstraction and Superstructuring for Sequence Classification. In: Proceedings of the IEEE Conference on Data Mining (ICDM 2009). IEEE press. Pp. 986-991

117. Towfic, F., Greenlee, H., and **Honavar, V**. (2009). Aligning Biomolecular Networks Using Modular Graph Kernels. In: Proceedings of the 9th Workshop on Algorithms in Bioinformatics (WABI 2009). Berlin: Springer-Verlag: LNBI Vol. 5724, pp. 345-361.

118. Towfic, F., Greenlee, H., and **Honavar, V**. (2009). Detecting Orthologous Genes Based on Protein-Protein Interaction Networks. In: Proceedings of the IEEE Conference on Bioinformatics and Biomedicine (BIBM 2009). IEEE Press. Pp. 48-53.

119. Yakhnenko, O., and **Honavar, V**. (2009). Multi-Modal Hierarchical Dirichlet Process Model for Predicting Image Annotation and Image-Object Label Correspondence. In: Proceedings of the SIAM Conference on Data Mining, SIAM. Pp. 281-294.

120. Bao, J., Voutsadakis, G., Slutzki, G., and **Honavar, V**. (2008). On the Decidability of Role Mappings between Modular Ontologies. Proceedings of the 23nd Conference on Artificial Intelligence (AAAI-2008), AAAI, 404-409.

121. Caragea D., Cook D., Wickham H. and **Honavar, V.** (2008). Visual Methods for Examining SVM Classifiers. Simeon J. Simoff, Michael H. Böhlen, Arturas Mazeika (Eds.): Visual Data Mining – Theory, Techniques and Tools for Visual Analytics. Springer-Verlag Lecture Notes in Computer Science Vol. 4404 pp.136-153

122. Caragea, C., Sinapov, J., Dobbs, D., and **Honavar, V.** (2008). Using global sequence similarity to enhance biological sequence labeling. Proceedings of the IEEE Conference on Bioinformatics and Biomedicine (BIBM-2008). IEEE Press. Pp. 104-111.

123. Dunn-Thomas, T., Dobbs, D.L., Sakaguchi, D. Young, M.J. **Honavar, V.** Greenlee, H. M. W. (2008). Proteomic Differentiation Between Murine Retinal and Brain Derived Progenitor Cells. Stem Cells and Development. 17: 191-131.

124. El-Manzalawy, Y., Dobbs, D., and **Honavar, V**. (2008). Predicting linear B-cell epitopes using string kernels. Journal of Molecular Recognition, DOI:10.1002/jmr.893. vol. 21. Pp. 243-255.

125. El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2008). On Evaluating MHC-II Binding Peptide Prediction Methods. PloS One, 3(9):e3268

126. El-Manzalawy, Y., Dobbs, D., and **Honavar, V**. (2008). Predicting Flexible Length Linear B-cell Epitopes, 7th International Conference on Computational Systems Bioinformatics, Stanford, CA. Singapore: World Scientific. Pp. 121–132.

127. El-Manzalawy, Y., Dobbs, D., and **Honavar, V.** (2008). Predicting Protective Linear B-cell Epitopes using Evolutionary Information. Proceedings of the IEEE Conference on Bioinformatics and Biomedicine (BIBM-2008). IEEE Press. Pp. 289-292

128. Hecker, L., Alcon, T., **Honavar, V**., and Greenlee, H. (2008). Analysis and Interpretation of Large-Scale Gene Expression Data Sets Using a Seed Network. Journal of Bioinformatics and Biology Insights. Vol. 2. Pp. 91-102.

129. Hughes, L., Bao. J., **Honavar, V**., and Reecy, J. (2008). Animal Trait Ontology (ATO): the importance and usefulness of a unified trait vocabulary for animal species. Journal of Animal Science 86: 1485-1491

130. Jo, H., Na, Y-C., Oh, B., Yang, J., and **Honavar, V**. (2008). Attribute Value Taxonomy Generation Using Matrix Based Genetic Algorithm. In: Proceedings of the IEEE Conference on Tools with Artificial Intelligence (ICTAI 2008). IEEE Press. Pp. 393-400.

131. Koul, N., Caragea, C., Bahirwani, V., Caragea, D., and **Honavar, V**. (2008). Using Sufficient Statistics to Learn Predictive Models from Massive Data Sets. Proceedings of the ACM/IEEE/WIC Conference on Web Intelligence (WI-2008). Pp. 923-926.

132. Koul, N., Lathrop, J., Lutz, J., and **Honavar, V.** (2008). Complexes of Online Self-Assembly. IEEE International Conference on Electro-Information Technology. IEEE Press. Pp. 448-452.

133. Lee, J-H., Hamilton, M., Gleeson, C., Caragea, C., Zaback, P., Sander, J.D., Li, X., Wu, F., Terribilini, M., **Honavar, V**., and Dobbs, D. (2008). Striking Similarities in Diverse Telomerase Proteins Revealed by Combining Structure Prediction and Machine Learning Approaches. Pacific Symposium on Bioinformatics, Vol. 13 pp. 501-512

134. Pathak, J., Basu, S., Lutz, R., and **Honavar, V.** (2008). MoSCoE: An Approach for Composing Web Services through Iterative Reformulation of Functional Specifications. International Journal on Artificial Intelligence Tools. Vol. 17. No. 1. Pp. 109-138.

135. Pathak, J., Basu, S., and **Honavar, V.** (2008). Assembling Composite Web Services from Autonomous Components. In: Emerging Artificial Intelligence Applications in Computer Engineering. Soldatos, J. et al. (ed). IOS Press. In press.

136. Pathak, J., Basu, S., and Honavar, V. (2008). Composing Web Services through Automatic Reformulation of Service Specifications. IEEE International Conference on Services Computing IEEE Press. Pp. 361-369.

137. Peto M., Kloczkowski A., Honavar V., Jernigan R.L. (2008). Use of machine learning algorithms to classify binary protein sequences as highly-designable or poorly-designable. BMC Bioinformatics, 9:487- **Highly Accessed.**

138. Santhanam, G., Basu, S., and **Honavar, V.** (2008). TCP-Compose* - A TCP-net based Algorithm for Efficient Composition of Web Services Based on Qualitative Preferences. Proceedings of the International Conference on Service-Oriented Computing. IEEE Press. Pp 453-467.

139. Tu, K. and **Honavar, V.** (2008). Unsupervised Learning of Probabilistic Context-Free Grammar using Iterative Biclustering. Proceedings of the International Colloquium on Grammatical Inference (ICGI-2008). Springer-Verlag Lecture Notes in Computer Science vol. 5278 pp. 224-237.

140. Voutsadakis, G., Bao, J., Slutzki, G., and **Honavar, V.** (2008). F-ALCI: A Fully Contextualized, Federated Logic for the Semantic Web. Proceedings of the ACM/IEEE/WIC Conference on Web Intelligence (WI-2008). Pp. 575-578.

141. Yan, C., Wu, F., Jernigan, R.L., Dobbs, D., **Honavar, V.** (2008) Analysis of Protein-Protein Interfaces. The Protein Journal 27 :59-70.

142. Andorf, C., Dobbs, D., and **Honavar, V.** (2007). Exploring Inconsistencies in Large-Scale Protein Function Annotations: A Machine Learning Approach. BMC Bioinformatics. http://www.biomedcentral.com/1471-2105/8/284/

143. Bao, J., Slutzki, G., and **Honavar, V.** (2007). A Semantic Importing Approach to Reusing Knowledge from Multiple Autonomous Ontology Modules. In: Proceedings of the 22nd Conference on Artificial Intelligence (AAAI-2007). Vancouver, Canada. AAAI Press. Pp. 1304- 1309

144. Bao, J., Slutzki, G. and **Honavar, V.** (2007). Privacy-Preserving Reasoning on the Semantic Web. Proceedings of the ACM/WIC/IEEE International Conference on Web Intelligence, IEEE Press, 791-797.

145. Caragea, C., Sinapov, J., Silvescu, A., Dobbs, D. And **Honavar, V.** (2007). Glycosylation Site Prediction Using Ensembles of Support Vector Machine Classifiers. BMC Bioinformatics. Doi:10.1186/1471-2105-8-438 **Highly Accessed.**

146. Caragea, C., Sinapov, J., Terribilini, M., Dobbs, D. and **Honavar, V.** (2007) Assessing the Performance of Macromolecular Sequence Classifiers. Proceedings of the IEEE Conference on Bioinformatics and Bioengineering, (IEEE BIBE 2007). IEEE Press. Pp. 320-326.

147. Helmer, G., Wong, J., Slagell, M., **Honavar, V**., Miller, L., Wang, Y., Wang, X., and Stakhanova, N. (2007). Software Fault Tree and Colored Petri Net Based Specification, Design, and Implementation of Agent-Based Intrusion Detection Systems. International Journal of Information and Computer Security. Vol. 1. No. 1/2. Pp. 109-142.

148. J. McCalley, **V. Honavar**, S. Ryan, W. Meeker, D. Qiao, R. Roberts, Y. Li, J. Pathak, M. Ye, Y. Hong (2007). Integrated Decision Algorithms for Auto-steered Electric Transmission System Asset Management. 7th Intl. Conference on Computational Science, Lecture notes in Computer Science. Berlin: Springer-Verlag. Vol. 4487. Pp. 1066-1073.

149. Pathak, J., Basu, S., and **Honavar, V**. (2007). On Context-Specific Substitutability of Web Services. Proceedings of the IEEE International Conference on Web Services, pp. 192-199. IEEE press.

150. Pathak, J., Li, Y., **Honavar, V.**, McCalley, J. (2007). A Service-Oriented Architecture for Electric Power Transmission System Asset Management. Second International Workshop on Engineering Service-Oriented Applications: Design and Composition, Lecture Notes in Computer Science, Berlin: Springer-Verlag, 2007.

151. Terribilini, M., Sander, J.D., Lee, J-H., Zaback, P., Jernigan, R.L., **Honavar, V**. and Dobbs, D. (2007). RNABindR: A Server for Analyzing and Predicting RNA Binding Sites in Proteins. Nucleic Acids Research. Vol. 35. No. 9. Pp. doi:10.1093/nar/gkm2, 2007

152. Wu, F., Towfic, F., Dobbs, D. and **Honavar, V**. (2007) Analysis of Protein Protein Dimeric Interfaces. IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007). San Jose, CA.

153. Bao, J., Caragea, D., and **Honavar, V**. (2006). A Tableau-based Federated Reasoning Algorithm for Modular Ontologies. ACM / IEEE / WIC Conference on Web Intelligence, Hong Kong, pp. 404-410.

154. Bao, J., Caragea, D., and **Honavar, V**. (2006). On the Semantics of Linking and Importing in Modular Ontologies. International Semantic Web Conference, Athens, Georgia, USA, Springer-Verlag Lecture Notes in Computer Science, Vol. 4273, pp. 72-86.

155. Bao. J. and **Honavar, V**. Representing and Reasoning with Modular Ontologies. In AAAI 2006 Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition (SweCKa 2006), Arlington, VA, USA, October 2006.

156. Bao. J., Caragea, D., and **Honavar, V.** Package-based Description Logics – Preliminary Results. International Semantic Web Conference (ISWC) 2006. Springer-Verlag Lecture Notes in Computer Science, Vol. 4273, pp. 967-969.

157. Bao, J., Caragea, D., and **Honavar, V.** (2006). Towards Collaborative Environments for Ontology Construction and Sharing. Proceedings of the International Symposium on Collaborative Technologies and Systems., Las Vegas, IEEE press. Pp. 99-108.

158. Bao, J., Hu, Z., Caragea, D., Reecy, J., and **Honavar, V.** (2006). A Tool for Collaborative Construction of Large Biological Ontologies. Fourth International Workshop on Biological Data Management (BIDM 2006), Krakov, Poland, IEEE Press. Pp. 191-195.

159. Bao, J., Caragea, D., and **Honavar, V.** (2006). Modular Ontologies – A Formal Investigation of Semantics and Expressivity. In Proceedings of the First Asian Semantic Web Conference, Beijing, China, Springer-Verlag Lecture Notes in Computer Science, Vol. 4185, pp. 616-631. **Best paper award**.

160. Bromberg, F., Margaritis, D., and **Honavar, V**. (2006). Efficient Markov Network Structure Discovery from Independence Tests. In: SIAM Conference on Data Mining (SDM 06). SIAM Press. Pp. 141-152.

161. Caragea, D., Zhang, J., Pathak, J., and **Honavar, V.** Learning Classifiers from Distributed, Ontology-Extended Data Sources. Proceedings of the 8th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2006), Krakov, Poland, Lecture Notes in Computer Science. Berlin: Springer. Vol. 4081. Pp. 363-373.

162. Caragea, D. and **Honavar, V.** (2006). A General Strategy for Learning Classifiers from Semantically Heterogeneous, Distributed Data. In AAAI 2006 Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition (SweCKa 2006), Arlington, VA, USA, AAAI. Pp. 1- 8.

163. Kang, D-K., Silvescu, A. and **Honavar, V**. (2006). Recursive Naive Bayes Learner for Sequence Classification. In: Proceedings of the Tenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2006). Lecture Notes in Computer Science. Vol. 3918, Berlin: Springer-Verlag. Pp. 45-54

164. Lee, K., Joo, J., Yang, J., and **Honavar, V**. (2006). Proceedings of the International Conference on Advanced Data Mining and Applications, Second International Conference (ADMA 2006), Springer-Verlag Lecture Notes in Computer Science. Vol. 4093. Pp. 465-472.

165. McCalley, J.D., **Honavar, V**., Ryan, S.M., Meeker, W.Q., Roberts, R.A., Qiao, D., and Li, Y.: Auto-steered Information-Decision Processes for Electric System Asset Management. In: Proceedings of the International Conference on Computational Science, pp. 440-447

166. Pathak, J., Basu, S., Lutz, R., and **Honavar, V.** (2006). Selecting and Composing Web Services Through Iterative Reformulation of Functional Specifications. Proceedings of the IEEE International Conference on Tools With Artificial Intelligence (ICTAI 2006), Washington, DC, IEEE. **Best Paper Award.**

167. Pathak, J., Basu, S., Lutz, R., and Honavar, V. (2006). Parallel Web Service Composition in MoSCoE: A Choreography Based Approach. Proceedings of the IEEE European Conference on Web Services (ECOWS 2006), Zurich, Switzerland, IEEE. Pp. 3-12.

168. Pathak, J., Basu, S., and **Honavar, V.** (2006). Modeling Web Services by Iterative Reformulation of Functional and Non-Functional Requirements. Proceedings of the International Conference on Service Oriented Computing. Springer-Verlag Lecture Notes in Computer Science. Vol. 4294. Pp. 314-326.

169. Pathak, J. Jiang, Y., **Honavar, V**., McCalley, J. (2006). Condition Data Aggregation for Failure Mode Estimation of Power Transformers. In IEEE 39th Hawaii International Conference on System Sciences (HICSS-2006), IEEE press, Kauai, Hawaii, USA.

170. Silvescu, A. and **Honavar, V**. (2006) Independence, Decomposability and functions which take values into an Abelian Group. Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics, 2006.

171. Terribilini, M., Lee, J-H., Yan, C., **Honavar, V**. and Dobbs, D. (2006). Identifying Interactions in ”Recalcitrant” Proteins. Predicted Protein and RNA binding sites in REV proteins of HIV and EIAV agree with experimental data. In: Proceedings of the Pacific Symposium on Biocomputing, Hawaii, World Scientific. Vol. 11. Pp. 415-426.

172. Vasile, F., Silvescu, A., Kang, D-K., and **Honavar, V**. (2006). TRIPPER: An Attribute Value Taxonomy Guided Rule Learner. In: Proceedings of the Tenth Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD). Lecture Notes in Computer Science. Vol. 3918, Berlin: Springer-Verlag. Pp. 55-59.

173. Wu, F., Olson, B., Dobbs, D., and **Honavar, V.** (2006). Using Kernel Methods to Predict Protein-Protein Interaction Sites from Sequence. IEEE Joint Conference on Neural Networks, Vancouver, Canada, IEEE Press. 2006.

174. Terribilini, M., Lee, J.H., Yan, C., Jernigan, R., **Honavar, V.,** and Dobbs, D. (2006). Computational Prediction of Protein-RNA Interfaces. RNA Journal.. Vol. 12. No. 1450. Pp. 1462, 2006.

175. Yan, C., Terribilini, M., Wu, F., Jernigan, R.L., Dobbs, D. and **Honavar, V**. (2006). Identifying amino acid residues involved in protein-DNA interactions from sequence. BMC Bioinformatics. Vol. 7. Pp. 262-, 2006. **Highly accessed.**

176. Wang, Y., Behera, S., Wong, J., Helmer, G., **Honavar, V**., Miller, L. and Lutz, R. (2006). Towards the automatic generation of mobile agents for distributed intrusion detection systems. Journal of Systems and Software. Vol. 79. Pp. 1-14.

177. Zhang, J., Silvescu, A., Kang, D-K., and **Honavar, V.** (2006). Learning Compact and Accurate Naive Bayes Classifiers from Attribute Value Taxonomies and Partially Specified Data. Knowledge and Information Systems. Vol. 9. No. 2. Pp. 157-179, 2006.

178. Caragea, D., Zhang, J., Bao, J., Pathak, J., and **Honavar, V**. (2005). Algorithms and Software for Collaborative Discovery from Autonomous, Semantically Heterogeneous, Distributed, Information Sources. Invited paper. In: Proceedings of the Conference on Algorithmic Learning Theory. Lecture Notes in Computer Science. Vol. 3734. Berlin: Springer-Verlag. Pp. 13-44.

179. Caragea, D., Silvescu, A., Pathak, J., Bao, J., Andorf, C., Dobbs, D., and **Honavar, V.** (2005) Information Integration and Knowledge Acquisition from Semantically Heterogeneous Biological Data Sources. Data Integration in Life Sciences (DILS 2005), San Diego, Berlin: Lecture Notes in Computer Science. Berlin: Springer-Verlag. Vol. 3615 pp. 175-190.

180. Caragea, D., Bao, J., Pathak, J., Andorf, C, Dobbs, D., and **Honavar, V.** (2005). Information Integration from Semantically Heterogeneous Biological Data Sources. Proceedings of the Third International Workshop on Biological Data Management (BIDM 2005), 16th International Workshop on Database and Expert Systems Applications (DEXA 2005), Copenhagen, Denmark. IEEE Computer Society 2005. Pp. 580-584

181. Kang, D-K., Silvescu, A., and **Honavar, V**. (2005). Multinomial Event Model Based Abstraction for Sequence and Text Classification. Proceedings of the Symposium on Abstraction, Reformulation, and Approximation (SARA 2005), Edinburgh, UK, Lecture Notes in Computer Science, Vol. 3607. Berlin: Springer-Verlag, pp. 134-148

182. Kang, D-K., Fuller, D., and **Honavar, V.** (2005). Learning Classifiers for Misuse and Anomaly Detection Using a Bag of System Calls Representation. Proceedings of the 6th IEEE Systems, Man, and Cybernetics Workshop (IAW 05), West Point, NY, IEEE Press, pp. 118-125

183. Kang, D-K., Fuller, D., and **Honavar, V**. Learning Misuse and Anomaly Detectors from System Call Frequency Vector Representation IEEE International Conference on Intelligence and Security Informatics, 2005. Lecture Notes in Computer Science, Vol. 3495. Berlin: Springer- Verlag. Pp. 511-516.

184. Pathak, J,, Koul, N., Caragea, D., and **Honavar, V.** (2005). A Framework for Semantic Web Services Discovery. In: Proceedings of the 7th ACM International Workshop on Web Information and Data Management (WIDM 2005). ACM Press. Pp. 45-50.

185. Wu, F., Zhang, J., and **Honavar, V.** (2005). Learning Classifiers Using Hierarchically Structured Class Taxonomies. In: Proceedings of the Symposium on Abstraction, Reformulation, and Approximation (SARA 2005), Edinburgh, UK, Lecture Notes in Computer Science, Vol. 3607. Berlin: Springer-Verlag, pp. 313-320.

186. Yakhnenko, O., Silvescu, A., and **Honavar, V.** (2005). Discriminatively Trained Markov Model for Sequence Classification. In: Proceedings of the IEEE Conference on Data Mining (ICDM 2005). IEEE Press. Pp. 498-505.

187. Zhang, J., Caragea, D. and **Honavar, V.** (2005). Learning Ontology-Aware Classifiers. In: Proceedings of the 8th International Conference on Discovery Science. Springer-Verlag Lecture Notes in Computer Science. Singapore. Vol. 3735. Pp. 308-321. Berlin: Springer-Verlag.

188. Andorf, C., Silvescu, A., Dobbs, D. and **Honavar, V.** Learning Classifiers for Assigning Protein Sequences to Gene Ontology Functional Families. In: Proceedings of the Fifth International Conference on Knowledge Based Computer Systems (KBCS 2004), India.

189. Bao, J., Cao, Y., Tavanapong, W., and **Honavar, V.** (2004). Integration of Domain-Specific and Domain-Independent Ontologies for Colonoscopy Video Database Annotation. In: International Conference on Information and Knowledge Engineering (IKE 04).

190. Caragea, D., Pathak, J., and **Honavar, V.** (2004). Learning Classifiers from Semantically Heterogeneous Data. In: Proceedings of the International Conference on Ontologies, Databases, and Applications of Semantics (ODBASE 2004), Springer-Verlag Lecture Notes in Computer Science vol. 3291. Pp. 963-980.

191. Caragea, D., Silvescu, A., and **Honavar, V.** (2004). A Framework for Learning from Distributed Data Using Sufficient Statistics and its Application to Learning Decision Trees. International Journal of Hybrid Intelligent Systems. **Invited** Paper. Vol 1. Pp. 80-89

192. Cook, D., Caragea, D. and **Honavar, V.** (2004). Visualization in Classification Problems. In: Proceedings in Computational Statistics (COMPSTAT 2004). Pp. 823-830.

193. Kang, D-K., Silvescu, A., Zhang, J., and Honavar, V. (2004). Generation of Attribute Value Taxonomies from Data for Data-Driven Construction of Accurate and Compact Classifiers. In: Proceedings of the IEEE International Conference on Data Mining. IEEE Press. Pp. 130-137.

194. Lonosky, P., Zhang, X., **Honavar, V**., Dobbs, D., Fu, A., and Rodermel, S. (2004) A Proteomic Analysis of Chloroplast Biogenesis in Maize. Journal of Plant Physiology. Vol. 134, pp. 560-574, 2004.

195. Pathak, J., Caragea, D., and **Honavar, V.** (2004). Ontology-Extended Component-Based Workflows: A Framework for Constructing Complex Workflows from Semantically Heterogeneous Software Components. In: Proceedings of the Workshop on Semantic Web and Databases (SWDB 2004). Springer-Verlag Lecture Notes in Computer Science Vol. 3372, pp. 41–56.

196. Polikar, R., Udpa, L., Udpa, S., and **Honavar, V.** (2004). An Incremental Learning Algorithm with Confidence Estimation for Automated Identification of NDE Signals. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. 51: 990-1001.

197. Sen,T.Z., Kloczkowski, A., Jernigan, R.L., Yan, C., Honavar, V., Ho, K-M., Wang, C-Z., Ihm, Y., Cao, H., Gu, X., and Dobbs, D. (2004). Predicting Binding Sites of Protease-Inhibitor Complexes by Combining Multiple Methods. BMC Bioinformtics. 5:205- **Highly Accessed**.

198. Yan, C., Dobbs, D. and **Honavar, V**. (2004). Identifying Protein-Protein Interaction Sites from Surface Residues – A Support Vector Machine Approach. Neural Computing Applications.13: 123-129.

199. Yan, C., Dobbs, D., and **Honavar, V.** (2004). A Two-Stage Classifier for Identification of Protein-Protein Interface Residues. Bioinformatics. 20. Pp. i371-378, 2004. (also In: Proceedings of the Conference on Intelligent Systems in Molecular Biology (ISMB) 2004.)

200. Zhang, J., and **Honavar, V**. (2004). Learning Naïve Bayes Classifiers from Attribute Value Taxonomies and Partially Specified Data. In: Proceedings of the Conference on Intelligent Systems Design and Applications (ISDA 2004).

201. Zhang, J. and **Honavar, V**. (2004). AVT-NBL – An Algorithm for Learning Compact and Accurate Naive Bayes Classifiers from Attribute Value Taxonomies and Data. In: Proceedings of the IEEE International Conference on Data Mining. IEEE Press. Pp. 289-296

202. Atramentov, A., Leiva, H., and **Honavar, V**. (2003). A Multi-Relational Decision Tree Learning Algorithm – Implementation and Experiments. In: Proceedings of the Thirteenth International Conference on Inductive Logic Programming. (Vol. 2835, Lecture Notes in Computer Science) Berlin: Springer-Verlag. Pp. 38-56.

203. Caragea, D., Cook. D., and **Honavar, V.** Towards Simple, Easy-to-Understand, yet Accurate Classifiers. In: Proceedings of the IEEE International Conference on Data Mining (ICDM 2003). Melbourne, Florida, 2003.

204. Caragea, D., Cook, D., and **Honavar. V.** Visualizing Ensemble of Hyperplane Classifiers. In: Proceedings of Visual Data Mining Workshop. IEEE International Conference on Data Mining (ICDM 2003). Melbourne, Florida, 2003. Springer Verlag.

205. Caragea, D., Silvescu, A., and **Honavar, V.** (2003) Decision Tree Induction from Distributed, Heterogeneous, Autonomous Data Sources. In: Proceedings of the Conference on Intelligent Systems Design and Applications (ISDA 2003). Springer Verlag.

206. Helmer, G., Wong, J., **Honavar, V.**, and Miller, L. (2003). Lightweight Agents for Intrusion Detection. Journal of Systems and Software. Vol. 67, No. 2. Pp. 109-122.

207. Reinoso-Castillo, J., Silvescu, A., Caragea, D., Pathak, J., and **Honavar, V.** (2003). An Ontology-Based Query-Centric Approach to Information Extraction and Integration from Heterogeneous, Distributed, Autonomous Data Sources. IEEE International Conference on Information Reuse and Integration.

208. Wang, X., Schroeder, D., Dobbs, D., and **Honavar, V**. (2003). Data-Driven Discovery of Rules for Protein Function Classification Based on Sequence Motifs. Information Sciences. Vol. 155. No. 1., pp. 1-18

209. Yan, C., Dobbs, D., and **Honavar, V**. (2003) Identification of Residues Involved in Protein- Protein Interaction from Amino Acid Sequence – A Support Vector Machine Approach. In: Proceedings of the Conference on Intelligent Systems Design and Applications (ISDA 2003). Springer-Verlag.

210. Zhang, J., and **Honavar, V**. (2003). Learning from Attribute-Value Taxonomies and Partially Specified Instances. In: Proceedings of the International Conference on Machine Learning. Washington, DC. AAAI Press. Pp. 880-887.

211. Z. Zhong, V. Vishwanathan, J. McCalley, **V. Honavar**, Multiagent System Solutions for Distributed Computing, Communications, and Data Integration Needs in the Power Industry. In. Proc. Of the 2003 IEEE PES Summer Meeting, Toronto, July, 2003.

212. Zhang, J., Silvescu, A., and **Honavar, V.** Ontology-Driven Induction of Decision Trees at Multiple Levels of Abstraction. In: Proceedings of the Symposium on Abstraction, Reformulation, and Approximation (SARA-2002). Kananaskis, Alberta, Canada. Lecture Notes in Computer Science. Berlin: Springer-Verlag. Pp. 316-323

213. Andorf, C., Dobbs, D., and **Honavar, V.** (2002). Discovering Protein Function Classification Rules from Reduced Alphabet Representations of Protein Sequences. In: Proceedings of the Conference on Computational Biology and Genome Informatics. Durham, North Carolina.

214. Helmer, G., Wong, J., Slagell, M., **Honavar, V**., Miller, L., and Lutz, R. (2002) A Software Fault Tree Approach to Requirements Specification of an Intrusion Detection System. Requirements Engineering. Vol 7 (4) (2002) pp. 207-220.

215. Helmer, G., Wong, J., **Honavar, V.**, and Miller, L. (2002). Automated Discovery of Concise Predictive Rules for Intrusion Detection. Journal of Systems and Software. Vol. 60. No. 3. Pp. 165-175.

216. Wang, X., Schroeder, D., Dobbs, D., and **Honavar, V**. (2002). Data-Driven Discovery of Protein Function Classifiers: Decision Trees Based on MEME Motifs Outperform Those Based on PROSITE Patterns and Profiles on Peptidase Families. In: Proceedings of the Atlantic Symposium on Computational Biology, Genome Information Systems & Technology.

217. Caragea, D., Cook, D., and **Honavar, V**. (2001). Gaining Insights into Support Vector Machine Classifiers Using Projection-Based Tour Methods. In: Proceedings of the Conference on Knowledge Discovery and Data Mining.

218. Helmer, G., Wong, J., Slagell, M., **Honavar, V**., Miller, L. and Lutz, R. (2001). A Software Fault Tree Approach to Requirements Analysis of an Intrusion Detection System. In: Proceedings of the Symposium on Requirements Engineering for Information Security, Indianapolis, IN, USA.

219. Mikler, A., **Honavar, V**. and Wong, J. (2001). Autonomous Agents for Coordinated Distributed Parameterized Heuristic Routing in Large Communication Networks. Journal of Systems and Software. Vol. 56. Pp. 231-246.

220. Parekh, R. and **Honavar, V.** (2001). Learning DFA from Simple Examples. Machine Learning. Vol. 44. Pp. 9-35.

221. Polikar, R., Udpa, L., Udpa, S., and **Honavar, V**. (2001) Learn++: An Incremental Learning Algorithm for Multi-Layer Perceptron Networks. IEEE Transactions on Systems, Man, and Cybernetics. Vol. 31. No. 4. Pp. 491-508.

222. Polikar, R., Shinar, R., **Honavar, V**., Udpa, L., and Porter, M. (2001). Detection and Identification of Odorants Using an Electronic Nose. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing.

223. Silvescu, A., and **Honavar, V**. (2001). Temporal Boolean Network Models of Genetic Networks and Their Inference from Gene Expression Time Series. Complex Systems. Vol. 13. Pp. 54-75. (preliminary version In: Proceedings of the Atlantic Symposium on Computational Biology, Genome Information Systems & Technology).

224. Wang, D., Wang, X., **Honavar, V**., and Dobbs, D. (2001). Data-Driven Generation of Decision Trees for Motif-Based Assignment of Protein Sequences to Functional Families. In: Proceedings of the Atlantic Symposium on Computational Biology, Genome Information Systems & Technology.

225. Wong, J., Helmer, G., Naganathan, V. Polavarapu, S., **Honavar, V.,** and Miller, L. (2001) SMART Mobile Agent Facility. Journal of Systems and Software. Vol. 56. Pp. 9-22.

226. Balakrishnan, K., Bousquet, O. & Honavar, V. (2000). Spatial Learning and Localization in Animals: A Computational Model and its Applications for Mobile Robots. Adaptive Behavior. Vol. 7. No. 2. Pp. 173-216.

227. Pai, P., Miller, L., Nilakanta, S., **Honavar, V.,** and Wong, J. (2000). In: Challenges of Information Technology Management in the 21st Century. Proceedings of the Eleventh International Conference of the Information Resources Management Association, Anchorage, Alaska. Challenges of Information Technology Management in the 21st Century M. Khosrowpour (Ed.) pp. 325-329.

228. Parekh, R., Yang, J. & **Honavar, V.** (2000). Constructive Neural Network Learning Algorithms for Multi-Category Pattern Classification. IEEE Transactions on Neural Networks. Vol. 11. No. 2. Pp. 436-451.

229. Parekh, R. and **Honavar, V**. (2000). On the Relationships between Models of Learning in Helpful Environments. In: Proceedings of the Fifth International Conference on Grammatical Inference. Lecture Notes in Artificial Intelligence Vol. 1891. Berlin: Springer-Verlag. Pp. 207- 220.

230. Polikar, R., Udpa, L., Udpa, S., and **Honavar, V.** (2000). Learn++: An Incremental Learning Algorithm for Multilayer Perceptron Networks. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP) 2000. Istanbul, Turkey.

231. Yang, J., Parekh, R. and **Honavar, V**. (2000). Comparison of Performance of Variants of Single-Layer Perceptron Algorithms on Non-Separable Data. Neural, Parallel, and Scientific Computations. Vol. 8. Pp. 415-438.

232. Bhatt, R., Balakrishnan, K., and Honavar, V. (1999). A Hybrid Model for Rodent Spatial Learning and Localization. In: Proceedings of the International Joint Conference on Neural Networks. Washington, D.C.

233. Chen, C-H., & **Honavar, V.** (1999). A Neural Network Architecture for Syntax Analysis. IEEE Transactions on Neural Networks. Vol. 10, No. 1. Pp. 94-114.

234. Dandu, R., L. Miller, S. Nilakanta, and V. **Honavar.** (1999). Populating a data warehouse with mobile agents. In: Proceedings of the Tenth International Conference of the Information Resources Management Association. Hershey, PA. pp. 468-473.

235. Janakiraman, J. & **Honavar, V.** (1999). Adaptive Learning Rate for Speeding up Gradient Descent Learning. Microcomputer Applications. Vol. 18. Pp. 89-95.

236. Parekh, R. and **Honavar, V.** (1999). Simple DFA are Polynomially Probably Exactly Learnable from Simple Examples. In: Proceedings of the International Conference on Machine Learning. Bled, Slovenia. San Mateo, CA: Morgan Kaufmann. Pp. 298-306.

237. Yang, J., Parekh, R., **Honavar, V.,** and Dobbs, D. (1999). Data-Driven Theory Refinement Using KBDistAl. In: Proceedings of the Conference on Intelligent Data Analysis (IDA 99). Berlin: Springer-Verlag. Pp. 331-342.

238. Yang, J., Parekh, R., Honavar, V., and Dobbs, D. (1999). Data-Driven Theory Refienement Algorithms for Bioinformatics. In: Proceedings of the International Joint Conference on Neural Networks. Washington, D.C. pp. 4064-4068.

239. Yang, J., Parekh, R., and **Honavar, V**. (1999). DistAl: An Inter-pattern Distance Based Constructive Learning Algorithm. Intelligent Data Analysis. Vol. 3. Pp. 55-73.

240. Balakrishnan, K., Bhatt, R., and **Honavar, V.** (1998). A Computational Model of Rodent Spatial Learning and Some Behavioral Experiments. In: Proceedings of the Twentieth Annual Meeting of the Cognitive Science Society. Madison, WI. Mahwah, NJ: Lawrence Erlbaum. Pp. 102-107.

241. Balakrishnan, K., Bhatt, R., and **Honavar, V.** (1998). Spatial Learning and Localization in Animals: A Computational Model and Behavioral Experiments. In: Proceedings of the Second European Conference on Cognitive Modeling. Pp. 112-119. Nottingham University Press.

242. Balakrishnan, K. & **Honavar, V.** (1998). Intelligent Diagnosis Systems. Journal of Intelligent Systems. Vol 8. Pp. 239-290.

243. Bousquet, O., Balakrishnan, K. & **Honavar, V.** (1998). Is the Hippocampus a Kalman Filter? In: Proceedings of the Pacific Symposium on Biocomputing. Singapore: World Scientific. Pp. 655-666.

244. Helmer, G., Wong, J., **Honavar, V.,** and Miller, L. (1998) Intelligent Agents for Intrusion Detection and Countermeasures. In: Proceedings of the IEEE Information Technology Conference. Pp. 121-124.

245. **Honavar, V.**, Miller, L. & Wong, J. (1998). Distributed Knowledge Networks: Design, Implementation and Applications. In: Proceedings of the IEEE Information Technology Conference. Pp. 87-90.

246. **Honavar, V.**, Miller, L., Yang, J. & Wong, J. (1998). Intelligent Mobile Agents for Information Retrieval and Knowledge Discovery from Distributed Data and Knowledge Sources. In: Proceedings of the IEEE Information Technology Conference. Pp. 99-104.

247. Leavens, G., Baker, A., **Honavar, V.**, Lavalle, S., and Prabhu, G. (1998). Programming is Writing: Why Student Programs Must be Carefully Evaluated. Mathematics and Computer Education. Vol. 32, pp. 284-295.

248. Miller, L., **Honavar, V.**, and Wong, J. (1998). An Object-Oriented Data Warehouse for Seamless Information Access from Heterogeneous Distributed Data and Knowledge Sources. In: Proceedings of the IEEE Information Technology Conference. Pp. 27-32.

249. Parekh, R. & **Honavar, V.** (1998). Constructive Theory Refinement in Knowledge-Based Neural Networks. In: Proceedings of the International Joint Conference on Neural Networks. Anchorage, Alaska. Pp. 2318-2323.

250. Parekh, R., Nichitiu, C., & **Honavar, V.** (1998). A Polynomial Time Incremental Algorithm for Regular Grammar Inference. In: Proceedings of the Fourth International Colloquium on Grammatical Inference*.* Lecture Notes in Artificial Intelligence Vol. 1433. Berlin: Springer- Verlag. Pp. 37-49.

251. Spartz, R. & **Honavar, V.** (1998). An Empirical Analysis of the Expected Source Values Rule. Microcomputer Applications Vol. 17, pp. 29-34.

252. Yang, J. & **Honavar, V.** (1998). DistAl: An Inter-Pattern Distance-Based Constructive Neural Network Learning Algorithm. In: In: Proceedings of the International Joint Conference on Neural Networks*.* Anchorage, Alaska. Pp. 2208-2213.

253. Yang, J., Pai, P., Miller, L., & **Honavar, V.** (1998). Mobile Intelligent Agents for Document Classification and Retrieval: A Machine Learning Approach. In: Proceedings of the European Symposium on Cybernetics and Systems Research. Berlin: Springer-Verlag. Pp. 707-712.

254. Yang, J., Havaldar, R., **Honavar, V**., Miller, L., & Wong, J. (1998). Coordination of Distributed Knowledge Networks Using the Contract Net Protocol. In: Proceedings of the IEEE Information Technology Conference*.* Pp. 71-74.

255. Yang, J. & Honavar, V. (1998). Feature Subset Selection Using a Genetic Algorithm. **Invited** paper. IEEE Intelligent Systems vol. 13, no. 2., pp. 44-49. **Among the 10 most cited articles in IEEE Intelligent Systems.**

256. Yang, J. & **Honavar, V.** (1998). Experiments with the Cascade-Correlation Algorithm. Microcomputer Applications. Vol. 17, pp. 40-46.

242. Balakrishnan, K. & **Honavar, V**. (1997). Spatial Learning for Robot Localization. In: Proceedings of International Conference on Genetic Programming*.* Stanford, CA. Morgan Kaufmann. Pp. 389-397.

243. Mikler, A., Wong, J. & **Honavar, V**. (1997). An Object-Oriented Approach to Modelling and Simulation of Routing in Large Communication Networks. Journal of Systems and Software Vol. 40, pp 151-164.

244. Mikler, A., Wong, J. & **Honavar, V.** (1997). Quo Vadis – A Framework for Intelligent Routing in Large High Speed Communication Networks. Journal of Systems and Software. 37 61-73.

245. Miller, L., **Honavar, V.** & Barta, T.A. (1997). Warehousing Structured and Unstructured Data for Data Mining. In: Proceedings of the American Society for Information Science Annual Meeting (ASIS 97). Washington, D.C.

246. Parekh, R.G. & **Honavar, V.** (1997) Learning DFA from Simple Examples. In Proceedings of the International Workshop on Algorithmic Learning Theory. Berlin: Springer-Verlag. Lecture notes in Computer Science vol. 1316. Pp. 116-131.

247. Parekh, R.G., Yang, J. & **Honavar, V.** (1997). MUPStart – A Constructive Neural Network Learning Algorithm for Multi-Category Pattern Classification. In: Proceedings of IEEE International Conference on Neural Networks*.* Houston, TX. Pp 1924-1929.

248. Parekh, R.G., Yang, J. & **Honavar, V.** (1997). Pruning Strategies for Constructive Neural Network Learning Algorithms. In: Proceedings of IEEE International Conference on Neural Networks. Houston, TX. Pp. 1960-1965.

249. Yang, J. & **Honavar, V.** (1997). Feature Subset Selection Using a Genetic Algorithm. In: Proceedings of International Conference on Genetic Programming. Stanford, CA. Morgan Kaufmann. Pp. 380-385.

250. Zhou, G., McCalley, J., & **Honavar, V.** (1997). Power System Security Margin Prediction Using Radial Basis Function Networks. In: Proceedings of the North American Power Symposium, Laramee, WY.

251. Balakrishnan, K. & **Honavar, V.** (1996). Analysis of Neurocontrollers Designed by Simulated Evolution. In: Proceedings of the IEEE International Conference on Neural Networks. Washington, D.C.

252. Balakrishnan, K. & **Honavar, V.** (1996). On Sensor Evolution in Robotics. In: Proceedings of the Conference on Genetic Programming. Koza, Goldberg, Fogel, Riolo (Ed.) Cambridge, MA: MIT Press. Pp. 455-460.

253. Balakrishnan, K. & **Honavar, V.** (1996). Experiments in Evolutionary Neuro-robotics. In: Proceedings of the World Congress on Neural Networks. San Diego, CA. pp. 1025-1031.

254. Chen, C. & **Honavar, V**. (1996). A Neural Network Architecture for High-Speed Database Query Processing. Microcomputer Applications 15 7-13.

255. Mikler, A., **Honavar, V**. & Wong, J. (1996). Utility-Theoretic Heuristics for Intelligent Adaptive Routing in Large Communication Networks. In: Proceedings of the Fourth International Conference on Telecommunication Systems. Nashville, TN. Pp. 660-676.

256. Mikler, A., **Honavar, V.** & Wong, J. (1996). Analysis of Utility-Theoretic Heuristics for Intelligent Adaptive Routing. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence. Portland, Oregon. Vol. 1. Pp. 96-102.

257. Parekh, R. & **Honavar, V.** (1996). An Incremental Interactive Algorithm for Regular Grammar Inference. In: Proceedings of the Third International Colloquium on Grammar Inference (ICGI-96). Berlin: Springer-Verlag. Pp. 238-249.

258. Yang, J., Parekh, R. & **Honavar, V**. (1996). Mtiling: A Constructive Neural Network Learning Algorithm for Multi-Category Pattern Classification. In: Proceedings of the World Congress on Neural Networks. San Diego, CA. pp. 182-188.

259. Yang, J. & **Honavar, V.** (1996). A Simple Randomized Quantization Algorithm for Network Pattern Classifiers. In: Proceedings of the World Congress on Neural Networks. San Diego, CA. pp. 223-229.

260. Balakrishnan, K. & **Honavar, V**. (1995). Properties of Genetic Representations of Neural Architectures. Proceedings of the World Congress on Neural Networks. Washington, D.C. pp. 807-813.

261. Chen, C., Yang, J., Balakrishnan, K., Parekh, R. & **Honavar, V.** (1995). Analysis of Decision Boundaries Generated by Constructive Neural Network Learning Algorithms. In: Proceedings of the World Congress on Neural Networks. Washington, D.C. pp. 628-635.

262. Chen, C. & **Honavar, V**. (1995). A Neural Memory Architecture for Content as Well as Address-Based Storage and Recall: Theory and Applications. Connection Science. 7 293-312.

263. Mikler, A., Wong, J. & **Honavar, V**. (1995). Adaptive Heursistic Routing in Very Large High Speed Communication Networks Using Quo Vadis: Experimental Results. In: Proceedings of the Third International Conference on Telecommunication Systems, Modelling, and Analysis. Nashville, TN. Pp. 66-75.

264. Chen, C. & **Honavar, V.** (1994). Neural Network Automata. In: Proceedings of the World Congress on Neural Networks. San Diego, CA. pp. 470-477.

265. Mikler, A., Wong, J. & **Honavar, V.** (1994). Quo Vadis – A Framework for Intelligent Traffic Management. In: Proceedings of the International Conference on Intelligent Information Management Systems. Washington, D. C. pp. 25-28.

266. **Honavar, V.** & Uhr, L. (1993). Generative Learning Structures and Processes for Generalized Connectionist Networks. Information Sciences 70 75-108.

267. Janakiraman, J. & **Honavar, V.** (1993). Adaptive Learning Rate for Speeding up Learning in Backpropagation Networks. In: SPIE Conference on Artificial Neural Networks, Proceedings of the SPIE. Vol. 1966. Pp. 225-235. Bellingham, MA: SPIE.

268. Janakiraman, J. & **Honavar, V.** (1993). Adaptive Learning Rate for Increasing Learning in Backpropagation Networks. In: Proceedings of the World Congress on Neural Networks. Vol. IV. Pp. 378-381. Portland, Oregon.

269. Mikler, A., Wong, J. & **Honavar, V.** (1993). Quo Vadis – A Framework for Adaptive Routing in Very Large High-Speed Communication Networks. In: Alspector, J., Goodman, R. & Brown, T. X. (Ed.) Proceedings of the International Workshop on Applications of Neural Networks to Telecommunications. Pp. 196-202. Princeton, New Jersey. Lawrence Erlbaum.

270. Parekh, R. & **Honavar, V.** (1993). Efficient Learning of Regular Languages Using Teacher-Generated Positive Samples and Learner-Generated Queries. In: Proceedings of the Fifth UNB AI Symposium. Goldfarb, L. (Ed.) New Brunswick, Canada. Pp. 195-203.

271. Thambu, P., Honavar, V. & Barta, T. (1993). Knowledge-base Consistency Maintenance in an Evolving Intelligent Advisory System. In: Proceedings of FLAIRS-93. Pp. 116-120. Fort Lauderdale, Florida.

272. Yang, J. & **Honavar, V**. (1993). A Hierarchical Representation Scheme for Three-Dimensional Object Recognition and Description. In: Proceedings of FLAIRS-93. Pp. 194-197. Fort Lauderdale, Florida.

273. Balakrishnan, K. & **Honavar, V.** (1992). Faster Learning in Multi-Layer Networks by handling output-layer flat-spots. In: Proceedings of International Joint Conference on Neural Networks. Beijing, China. Vol. 2. Pp. 486-492.

274. Balakrishnan, K. & **Honavar, V.** (1992). Improving Convergence of Back-propagation by Handling Flat-Spots in the Output Layer. In: Proceedings of the Second International Conference on Artificial Neural Networks. Pp. 1003-1009. Brighton, UK.

275. **Honavar, V.** (1992). Some Biases for Efficient Learning of Spatial, Temporal, and Spatio-Temporal Patterns. In: Proceedings of International Joint Conference on Neural Networks. Beijing, China. Vol. 2. Pp. 419-27.

276. **Honavar, V.** (1992). Inductive Learning Using Generalized Distance Measures. SPIE Conference on Adaptive and Learning Systems. Proceedings of SPIE. Vol. 1706. Pp. 50-58. Bellingham, MA: SPIE.

277. Mikler, A., **Honavar, V**., & Wong, J. (1992). Simulating a Traveller: A Heuristic Approach to Routing in Large Communication Networks. In: Proceedings of the European Simulation Symposium. Pp. 297-301. Dresden, Germany.

278. Parekh R., Balakrishnan, K. & **Honavar, V**. (1992). Empirical Comparison of Flat-Spot Elimination Techniques in Back-propagation Networks. In: Proceedings of Simtec/WNN92*.* Houston, Texas. Pp. 463-468.

279. Spartz, R. & **Honavar, V.** (1992). An Empirical Analysis of the Expected Source Values Rule. In: Proceedings of International Joint Conference on Neural Networks. Beijing, China. Vol. 2. Pp. 480-485.

280. Yang, J. & **Honavar, V**. (1991). Experiments with the Cascade Correlation Algorithm. In: Proceedings of the Fourth UNB Artificial Intelligence Symposium. Fredericton, Canada. Pp. 369-380.

281. **Honavar, V**. & Uhr, L. (1990). Successive Refinement of Multi-Resolution Internal Representations of the Environment in Connectionist Networks. In: Proceedings of the Second Conference on Neural Networks and Parallel-Distributed Processing*.* Indiana University-Purdue University. Pp. 90-99.

282. **Honavar, V.** & Uhr, L. (1990). Coordination and Control Structures and Processes: Possibilities for Connectionist Networks. Journal of Experimental and Theoretical Artificial Intelligence 2 277-302.

283. **Honavar, V.** & Uhr, L. (1989). Brain-Structured Connectionist Networks that Perceive and Learn. Connection Science 1 139-159.

284. **Honavar, V.** & Uhr, L. (1989). Generation, Local Receptive Fields, and Global Convergence Improve Perceptual Learning in Connectionist Networks. In: Proceedings of the 1989 International Joint Conference on Artificial Intelligence*,* San Mateo, CA: Morgan Kaufmann. Pp. 180-185.

285. **Honavar, V.** & Uhr, L. (1988). A Network of Neuron-Like Units That Learns To Perceive By Generation As Well As Reweighting Of Its Links. In: Proceedings of the 1988 Connectionist Models Summer School*,* D. S. Touretzky, G. E. Hinton & T. J. Sejnowski (ed). Pp. 472-484. San Mateo, CA: Morgan Kaufmann.

Invited or Refereed Book Chapters

1. Tuggle, C., Towfic, F., and **Honavar, V**. (2011). An Introduction to Computational Systems Biology for Animal Scientists. In: Systems Biology For Animal Scientists. In: te Pas, M.F.W., Woelders, H., and Bannick, A. (Ed). Wiley.
2. Caragea, C. and **Honavar, V**. (2009). Machine Learning in Computational Biology. In: *Encyclopedia of Database Systems.* Raschid, L. (ed). Springer. Pp. 1663-1667
3. Caragea, D. and **Honavar, V**. (2009). Learning Classifiers from Distributed Data. In: *Encyclopedia of Database Technologies and Applications*, Ferraggine, V.E., Doorn, J.H., and Rivero, L.C. (Ed). New York: Idea Group.
4. **Honavar, V.** and Caragea, D. (2008). Towards a Semantics-Enabled Infrastructure for Knowledge Acquisition from Distributed Data. In: Next Generation Data Mining. Kargupta, H. et. Al., CRC Press. Pp. 317-337.

5. Caragea, D. and **Honavar, V**. (2008). *Learning Classifiers from Semantically Heterogeneous Data*. In: Encyclopedia of Data Warehousing and Mining. Wang, J. (Ed).
6. Pathak, J., Basu, S., and **Honavar, V.** (2008). Assembling Composite Web Services from Autonomous Components. In: Emerging Artificial Intelligence Applications in Computer Engineering. Soldatos, J. et al. (ed). IOS Press. Pp. 394-405.
7. **Honavar, V.** Miller, L., and Wong, J. (2004). Distributed Knowledge Networks. In: Unifying Themes in Complex Systems. Bar-Yam, Y. and Minai, A. (Ed). New York: Perseus Books.
8. McCalley, J., **Honavar, V.,** Zhang, Z., and Vishwanathan, V. (2004). Multiagent negotiation models for power system applications**, Invited Chapter.** In: Autonomous Systems and Intelligent Agents in Power System Control and Operation. Christian Rehtanz (Ed). 2004.
9. Pai, P., L.L. Miller, **V. Honavar,** J. Wong, and S. Nilakanta. (2002). Supporting Organizational Knowledge Management with Agents. In D. White (eds) Knowledge Mapping and Management. Pages 266-280.
10. Caragea, D., Silvescu, A., and **Honavar, V.** (2001). **Invited** Chapter. Towards a Theoretical Framework for Analysis and Synthesis of Agents That Learn from Distributed Dynamic Data Sources. In: Emerging Neural Architectures Based on Neuroscience*.* Berlin: Springer-Verlag. Pp. 547-559.
11. Balakrishnan, K. & **Honavar, V.** (2001). Evolving Neurocontrollers and Sensors for Artificial Agents. In: Evolutionary Synthesis of Intelligent Agents*.* Patel, M., Honavar, V. and Balakrishnan, K. (Ed). Cambridge, MA: MIT Press. Pp. 109-152. In press.
12. Chen, C. & Honavar, V. (2000). A Neural Network Architecture for Information Retrieval and Query Processing. **Invited** chapter In: Handbook of Natural Language Processing. Dale, Moisl & Somers (Ed). New York: Marcel Dekker. Pp. 873-888.
13. **Honavar, V.** and Balakrishnan, K. (2001). Evolutionary Synthesis of Neural Architectures for Intelligent Agents. In: Evolutionary Synthesis of Intelligent Agents*.* Patel, M., Honavar, V. & Balakrishnan, K. (Ed). Cambridge, MA: MIT Press. In press. Pp. 1-28.
14. Balakrishnan, K., and **Honavar, V.** (2000). Some Experiments in the Evolution of Robot Sensors. **Invited** chapter. In: Evolution of Engineering and Information Systems and Their Applications. Jain, L. (Ed). New York: CRC Press. Pp. 191-228.
15. Parekh, R. & **Honavar, V.** (2000). Automata Induction, Grammar Inference, and Language Acquisition. **Invited** chapter. In: Handbook of Natural Language Processing*.* Dale, Moisl & Somers (Ed). New York: Marcel Dekker. Pp. 727-764. In press.
16. **Honavar, V.,** Parekh, R., and Yang, J. (1998). Constructive Learning and Structural Learning. **Invited** article In: Encyclopedia of Electrical and Electronics Engineering*,* Webster, J. (Ed.), New York: Wiley. Vol. 4. Pp. 226-231.
17. **Honavar, V.** (1998). Machine Learning. **Invited** article In: *Encyclopedia of Electrical and Electronics Engineering*, Webster, J. (Ed.), New York: Wiley. Vol. 11. Pp. 656-659.
18. Yang, J. & **Honavar, V.** (1998). Feature Subset Selection Using A Genetic Algorithm. **Invited** chapter. In: Feature Extraction, Selection, and Construction: A Data Mining Perspective*.* Liu, H. & Motoda, H. (Ed.) New York: Kluwer. Pp. 117-136.
19. Honavar, V., & Uhr, L. (1995). Integrating Symbol Processing and Connectionist Networks. **Invited** chapter. In: Intelligent Hybrid Systems*.* Goonatilake, S. & Khebbal, S. (Ed). London: Wiley. Pp. 177-208.
20. **Honavar, V.** (1994). Symbolic Artificial Intelligence and Numeric Artificial Neural Networks: Toward a Resolution of the Dichotomy. **Invited** chapter In: Computational Architectures for Integrating Symbolic and Neural Processes. Sun, R. & Bookman, L. (Ed). New York: Kluwer. Pp. 351-388.
21. **Honavar, V.** (1994). Toward Learning Systems That Integrate Multiple Strategies and Representations. In: Artificial Intelligence and Neural Networks: Steps Toward Principled Integration*.* Honavar, V. & Uhr, L. (Ed). Pp. 615-644. New York: Academic Press.

22. Uhr, L. & **Honavar, V.** (1994). Artificial Intelligence and Neural Networks: Steps Toward Principled Integration. In: Artificial Intelligence and Neural Networks: Steps Toward Principled Integration. Honavar, V. & Uhr, L. (Ed). Pp. xvii-xxxii. New York: Academic Press.

Conference Proceedings Edited

1. **Honavar, V.** & Slutzki, G. (Ed.) (1998). **Grammatical Inference** Vol. 1433. Lecture Notes in Computer Science. Berlin: Springer-Verlag.
2. Banzaf, W., Daida, J., Eiben, A. Garzon, M., **Honavar, V**., Jakiela, M., & Smith, R. (Ed.) (1999). Proceedings of the Genetic and Evolutionary Computation Conference. San Mateo, CA: Morgan Kaufmann.
3. W. Langdon, E. Cantu-Paz, K. Mathias, R. Roy, D. Davis, R. Poli, K. Balakrishnan, **V. Honavar**, G. Rudolph, J. Wegener, L. Bull, M. Potter, A. Schultz, J. Miller, E. Burke, N. Jonoska. (2002). (Ed). Proceedings of the Genetic and Evolutionary Computing Conference. Palo Alto, CA: Morgan Kaufmann.
4. H. J. Caulfield, S.-H. Chen, H.-D. Cheng, R. Duro, **V. Honavar**, E. E. Kerre, M. Lu, M. G. (2002). Romay, T. K. Shih, D. Ventura, P. P. Wang, and Y. Yang, editors, 2002. Proceedings 6th Joint Conference on Information Sciences, JCIS / Association for Intelligent Machinery.

Plenary Lectures, Invited Keynotes Invited Conference Talks, Distinguished Lectures

1. Invited Talk, The Promise and Potential of AI, Machine Learning and Data Science in Integrating Data and Models Across Modalities, Scales, and Model Systems, NIH Workshop on Validation of Animal Models for Biomedical Research, January 2021.
2. Invited Plenary Talk, On the Importance of Causal Models in Making Sense of Big Data, Workshop on Brain, Computation, and Learning, Indian Institute of Science, 2019.
3. Invited Plenary Talk, An excursion into Network Neuroscience, Workshop on Brain, Computation, and Learning, Indian Institute of Science, 2018.
4. Invited Plenary Talk, Brain Sciences Meet Big Data. Workshop on Brain, Computation, and Learning, Indian Institute of Science, 2017
5. Invited Plenary Talk, Foundations of Data Science: Mind the Gaps! Workshop on Transdisciplinary Foundations of Data Science. Institute for Mathematics and Applications, University of Minnesota, 2016.
6. Plenary Talk, Accelerating Science: A Grand Challenge for AI? AAAI/CCC Symposium on Accelerating Science: A Grand Challenge for AI. 2016.
7. Invited Talk, Uncertainty in AI and Machine Learning. CCC Workshop on Uncertainty in Computation. 2014.
8. Invited Talk, Informatics Challenges in Understanding Interactions of Food, Energy, and Water Systems, NSF INFEWS Workshop, 2015.
9. Invited Keynote Talk, Computational Analysis and Prediction of Protein Interfaces and Interactions. 2014 Mid-south Computational Biology and Bioinformatics Conference, 2014

10. Invited Keynote Talk, From Big Data Analytics to Computational Discovery and Discovery Informatics. 9th IEEE International Conference on Collaborative Computing: Networking, Applications and Worksharing, Austin, Texas, 2013.

11. Invited Talk, From Data Analytics to Discovery Informatics. NIH NIAID Symposium on Unlocking the Power of Big Data. National Institutes of Health, 2013.

12. Invited Keynote Talk, From Big Data Analytics to Discovery Informatics. Conference on Complex Adaptive Systems, Washington DC. November 2012.

13. Invited Talk, Computational Prediction of Protein Interfaces and Interactions. Conference on Modeling Protein Interactions, Lawrence, Kansas, November 2012.

14. Invited Keynote Talk, Learning Predictive Models from Distributed Data. Conference on Intelligent Data Understanding, Boulder, CO, October 2012.

15. Invited Keynote Talk, Humanities as Information Sciences. Chicago Colloquium on Digital Humanities and Computer Science, Chicago, November 2009.

16. Invited Keynote Talk, Aligning Macromolecular Networks. Sixth International Biotechnology and Bioinformatics Symposium (BIOT 2009), Lincoln, Nebraska, October 2009.

17. Invited Plenary Talk, Machine Learning in Bioinformatics, Annual Conference of the Italian Association for Artificial Intelligence (AI*IA 2008), Cagliari, Italy, September 2008.

18. Invited Keynote Talk, International Congress on Pervasive Computing and Management (ICPCM 2008), New Delhi, India, December 2008.

19. Invited Talk, Telluride Meeting on Characterizing the Landscape From Biomolecules to Cellular Networks, Telluride, Colorado, July 2008.

20. Invited Talk, Privacy-preserving Reasoning, Semantic Technology Conference, San Jose, CA, USA, May 2008

21. Invited Keynote Talk, Computational Structural Bioinformatics Workshop, IEEE Conference on Bioinformatics and Biomedicine, Silicon Valley, 2007.

22. Invited Talk, Making Biology and Medicine a Predictive Science. NSF Workshop on Biomedical Informatics. Oregon, 2007.

23. Invited Talk, Knowledge Acquisition from Semantically Disparate Distributed Data. NSF Workshop on Next Generation Data Mining and Cyber-Enabled Discovery, Baltimore, Maryland, 2007.

24. Invited Talk, On Selective Sharing and Reuse of Ontologies, Semantic Technology Conference, San Jose, CA, USA, May 2007

25. Invited Keynote Talk, Semantic Web for Collaborative e-Science, International Conference on Intelligent Sensing and Information Processing, Bangalore, India, 2006.

26. Invited Lecture: Querying Semantically Heterogeneous Data Sources from a User's Point of View, Semantic Technology Conference, San Jose, CA, USA, March 2006.

27. Invited Plenary Talk, Algorithms and Software for Collaborative Discovery from Semantically Heterogeneous, Distributed, Autonomous Information Sources Sixteenth International Conference on Algorithmic Learning Theory (ALT 2005) and Eighth International Conference on Discovery Science (DS 2005). Singapore.

28. Plenary Talk, Data-Driven Discovery of Macromolecular Sequence-Structure-Function Relationships. International Conference on Intelligent System Design and Applications, 2003.

29. Invited Talk, Agent-Based Distributed Intelligent Information Networks for Computational Inference and Knowledge Discovery in Bioinformatics. In: Workshop on Agents in Bioinformatics, Italy, 2002.

30. Plenary Talk, Computational Discovery of Protein Sequence-Structure-Function Relationships, Diversity in Information Science and Technology, Nebraska EPSCOR Conference, 2002

31. Invited Keynote Talk, Learning from Large, Distributed, Heterogeneous Data Sets. International Symposium on Artificial Intelligence (ISAI 2001), Kolhapur, India.

32. Invited Talk, Distributed Intelligent Information Networks. Midwestern Conference on Artificial Intelligence and Cognitive Science, 2000.

33. Invited Talk, Cumulative Learning in Open Environments. International Workshop on Current Computational Architectures Integrating Neural Networks and Neuroscience. Durham Castle, United Kingdom. 2000.

34. Invited Talk, Distributed Knowledge Networks. Artificial Intelligence for Distributed Information Networks (AiDIN '99) Workshop held during the 1999 National Conference on Artificial Intelligence (AAAI 99), Orlando, Florida. July 1999.

Juried Papers, Lightly Refereed Papers, Extended Abstracts, and Posters in Conferences and Workshops

1. Santhanam, G., Basu, S., and Honavar, V. (2011). Identifying Sustainable Designs Using Preferences Over Sustainability Attributes. In: AAAI Spring Symposium on Artificial Intelligence in Sustainable Design. Stanford, CA.

2. Caragea, C., Silvescu, A., Caragea, D. and Honavar, V. (2009). Abstraction Augmented Markov Models. In: NIPS Workshop on Machine Learning in Computational Biology (MLCB).

3. Tu, K., and Honavar, V. (2009). An Empirical Study of Hierarchical Dirichlet Process Priors for Grammar Induction. In: NIPS Workshop on Grammar Induction, Representation of Language, and Language Learning.

4. Yakhnenko, O. and Honavar, V. (2009). Multiple label prediction for image annotation with multiple kernel correlation models. In: CVPR Workshop on Visual Context Learning.

5. Santhanam, G., Basu, S., and Honavar, V. (2008). On Utilizing Qualitative Preferences in Web Service Composition: A CP-net based approach 2nd International Workshop on Web Service Composition and Adaptation (WSCA) 2008

6. Towfic, F., Gemperline, D., Caragea, C., Wu, F., Dobbs, D., and Honavar, V. Structural Characterization of RNA-Binding Sites of Proteins: Preliminary Results. *IEEE BIBM Computational Structural Bioinformatics Workshop*, 2007.

7. Andorf, C., Dobbs, D., and Honavar, V. Potential Errors in Mouse Protein Gene Ontology Annotations Returned by AmiGO. Oral Presentation in: Gene Ontology Users Workshop, MGED, Seattle, Washington, September, 2006.

8. A Bao, J., Caragea, D., and Honavar, V. (2006). A Distributed Tableau Algorithm for Package-based Description Logics. Proceedings of the Second International Workshop on Context Representation and Reasoning (CRR 2006), Riva del Garda, Italy, CEUR. 2006.

9. Bao. J. and Honavar, V. (2006) Divide and Conquer Semantic Web with Modualr Ontologies – A Brief Review of Modular Ontology Language Proposals. First International Workshop on Modular Ontologies (WoMo2006). International Semantic Web Conference, Athens, GA, 2006.
10. Bao. J. and Honavar, V. (2006) Adapt OWL as a Modular Ontology Language (Position Paper). Accepted by OWL: Experiences and Directions (OWLED 2006). International Semantic Web Conference, Athens, GA. CEUR Report. Vol. 216.
11. Caragea, D. and Honavar, V. (2006). Knowledge Discovery from Disparate Earth Data Sources. Second NASA Data Mining Workshop: Issues and Applications in Earth Sciences. Poster Session. Pasadena, CA, May 23-24, 2006.
12. Pathak, J., Basu, S., and Honavar, V. (2006). Modeling Web Service Composition Using Symbolic Transition Systems. AAAI '06 Workshop on AI-Driven Technologies for Services- Oriented Computing (AI-SOC), Boston, MA, 2006.
13. J. Pathak, S. Basu, R. Lutz, and V. Honavar. (2006). MoSCoE: A Framework for Modeling Web Service Composition and Execution. IEEE Conference on Data Engineering Ph.D. Workshop, Atlanta, GA, 2006.
14. Sander, J., Fu, F. Terribilini, M., Townsend, J., Winfrey, R., Wright, D., Lee, J.J., Zaback, P., F. Wu, F., Honavar, V., Voytas, D. and Dobbs, D. (2006) Designing C2H2 Zinc Finger Proteins to Target Specific Sites in Genomic DNA. 10th Annual Pacific Symposium on Biocomputing (PSB 2006), Maui, Hawaii. Poster Presentation.
15. Andorf, C., Dobbs, D., and Honavar, V. (2006) Learning classifiers for assigning sequences to subcellular localization families. Intelligent Systems in Molecular Biology (ISMB 2006), Fortaleza, Brazil. Poster Presentation.
16. EL-Manzalawy, Y., Caragea, C., Dobbs, D., Honavar, V. (2006) On the quality of motifs for protein phosphorylation site prediction. Intelligent Systems in Molecular Biology (ISMB 2006), Fortaleza, Brazil. Poster Presentation.
17. Terribilini, M., Sander, J., Olson, B., Lee, J.-H., Jernigan, R., Honavar, V., and Dobbs, D. (2006) A computational method to identify amino acid residues involved in protein-RNA interactions. Intelligent Systems in Molecular Biology (ISMB 2006), Fortaleza, Brazil. Poster Presentation.
18. Andorf, C., Silvescu, A., Dobbs, D. and Honavar, V. (2005). Learning Classifiers for Assigning Proteins to Gene Ontology Functional Families. Poster Presentation. Intelligent Systems in Molecular Biology (ISMB 2005).
19. Bao, J. and Honavar, V. (2005). Collaborative Package-Based Ontology Building and Usage. In: IEEE Workshop on Knowledge Acquisition from Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources. Held in conjunction with the IEEE International Conference on Data Mining (ICDM 2005), Houston, Tx.
20. Caragea, C., Caragea, D., and Honavar, V. (2005). Learning Support Vector Machine Classifiers from Distributed Data. Proceedings of the 22nd National Conference on Artificial Intelligence (AAAI 2005).
21. Caragea, D., Silvescu, A., Bao, J., Pathak, J., Andorf, C., Yan, C., Dobbs, D., and Honavar, V. (2005) Poster presentation. Knowledge Acquisition from Semantically Heterogeneous, Autonomous, Distributed Data Sources. Intelligent Systems in Molecular Biology (ISMB 2005).

22. Terribilini, M., Yan, C., Lee, J-H, Honavar, V. and Dobbs, D. (2005). Computational Prediction of RNA binding sites in Proteins based on Amino Acid Sequence. Poster Presentation. Intelligent Systems in Molecular Biology (ISMB 2005).
23. Terribilini, M., Lee, J-H., Sen, T., Yan, C., Andorf, C., Sparks, W., Carpenter, S., Jernigan, R., Honavar, V., and Dobbs, D, (2005). Computational Identification of RNA binding sites in proteins. Poster presentation. Pacific Symposium on Biocomputing (PSB 2005).
24. Vasile, F., Silvescu, A., Kang, D-K., and Honavar, V. (2005): TRIPPER: Rule Learning Using Attribute Value Taxonomies. In: AAAI-05 Workshop on Human-Comprehensible Machine Learning.
25. Yan, C., Terribilini, M., Wu, F., Dobbs, D. and Honavar, V. (2005). A Computational Method for Identifying Amino Acid Residues Involved in Protein-DNA interactions. Poster Presentation. Intelligent Systems in Molecular Biology (ISMB 2005).
26. Bao, J. and Honavar, V. Collaborative Ontology Building with Wiki@nt. In: Proceedings of the Third International Workshop on Evaluation of Ontology Building Tools, http://km.aifb.uni-karlsruhe.de/ws/eon2004/EON2004_Proceedings.pdf. The Third International Semantic Web Conference (ISWC-0004), Hiroshima, 2004.
27. D.-K. Kang, A. Silvescu, J. Zhang, and V. Honavar, "Generation of Attribute Value Taxonomies from Data and Their Use in Data-Driven Construction of Accurate and Compact Naive Bayes Classifiers," Proceedings of ECML/PKDD-2004 Knowledge Discovery and Ontologies Workshop (KDO-2004), Pisa, Italy, September 24, 2004.
28. Silvescu A., and Honavar V. (2004). A Graphical Model for Shallow Parsing Sequences. In: The AAAI-04 Workshop on Adaptive Text Extraction and Mining (ATEM-2004). July 2004, San Jose CA.
29. Yan, C., Honavar, V. and Dobbs, D. (2004). Application of a Two-Stage Method for Identification of Protein-Protein Interface Residues. Poster Presentation. Eighth Annual International Conference on Research in Computational Molecular Biology (RECOMB 2004).
30. Caragea, D., Reinoso-Castillo, J., Silvescu, A., and Honavar, V. Statistics Gathering for Learning from Heterogeneous, Distributed, Autonomous Data Sources. In: Proceedings of the Workshop on Information Integration on the Web. International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 2003.
31. Atramentov, A., and Honavar, V. Speeding up Multi-Relational Data Mining. In: Proceedings of the Workshop on Learning Statistical Models from Relational Data. International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 2003.
32. Honavar, V., Dobbs, D., Jernigan, R., Caragea. D., Reinoso-Castillo, J., Silvescu, A., Pathak, J., Andorf, C., Yan, C., and Zhang, J. (2003). Algorithms and Software for Information Extraction, Integration, and Data-Driven Knowledge Acquisition from Heterogeneous, Distributed, Autonomous, Biological Information Sources. Poster Presentation. Biomedical Information Science and Technology Initiative (BISTI) Symposium; Digital Biology: The Emerging Paradigm. National Institutes of Health.
33. Silvescu A., and Honavar V. (2003) Ontology Elicitation: Structural Abstraction = Structuring + Abstraction + Multiple Ontologies. Poster presentation. Learning Workshop, Snowbird, Utah, 2003.

34. Leiva, H., Atramentov, A., and Honavar, V. (2002). Experiments with MRDTL – A Multirelational Decision Tree Learning Algorithm. In: Proceedings of the Workshop on Multi- Relational Decision Tree Learning. http://www-ai.ijs.si/SasoDzeroski/MRDM2002/

35. Silvescu, A., Reinoso-Castillo, J., Andorf, C., Honavar, V., and Dobbs, D. (2001). Ontology-Driven Information Extraction and Knowledge Acquisition from Heterogeneous, Distributed Biological Data Sources. In: Proceedings of the IJCAI-2001 Workshop on Knowledge Discovery from Heterogeneous, Distributed, Autonomous, Dynamic Data and Knowledge Sources.

36. Bhatt, R., Balakrishnan, K., and Honavar, V. (2000). Representation and Learning of Spatial Maps. In: Workshop on Machine Learning of Spatial Knowledge, International Conference on Machine Learning (ICML-2000), Stanford University.

37. Caragea, D., Silvescu, A., and Honavar, V. (2000). Distributed and Incremental Learning Using Extended Support Vector Machines. In: Proceedings of the 17th National Conference on Artificial Intelligence. Austin, TX.

38. Caragea, D., Silvescu, A., and Honavar, V. (2000). Multi-Agent Learning from Distributed Data Sources. In: Workshop on Multi-Agent learning: Theory and Practice International Conference on Machine Learning (ICML-2000), Stanford University.

39. Caragea, D., Silvescu, A., and Honavar, V. (2000). Agents that Learn from Distributed Dynamic Data Sources. In: Proceedings of the Workshop on Learning Agents, Agents 2000/ECML 2000. Stone, P. and Sen, S. (Ed.) ECML. Barcelona, Spain. Pp. 53-61.

40. Caragea, D., Silvescu, A., and Honavar, V. (2000). Distributed, Parallel, and Incremental Learning: A Theoretical Model. In: KDD 2000 Workshop on Distributed and Parallel Knowledge Discovery. Boston, MA.

41. Silvescu, A., and Honavar, V. (2000). Genetic Network Inference from Gene Expression Data. In: Workshop on Learning from Sequential and Temporal Data. International Conference on Machine Learning (ICML 2000), Stanford University.

42. Helmer, G., Wong, J., Honavar, V., and Miller, L. (1999). Data-Driven Induction of Compact Predictive Rules for Intrusion Detection from System Log Data. In: Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO 99). San Mateo, CA: Morgan Kaufmann. Pp. 1781.

43. Tiyyagura, A., Chen, F., Yang, J., and Honavar, V. (1999). Feature Subset Selection in Rule Induction. In: Proceedings of the Conference on Genetic and Evolutionary Computation (GECCO 99). San Mateo, CA: Morgan Kaufmann. Pp. 1800.

44. Honavar, V. (1997). The Design Process: A Computational Perspective. In: Proceedings of the NSF Workshop on Decision-based Design, Sacramento, CA.

45. Balakrishnan, K. & Honavar, V. (1996). Experiments in Evolutionary Synthesis of Neurocontrollers. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence. AAAI Press. Vol. 2. Pp. 1378.

46. Parekh, R. & Honavar, V. (1996). An Incremental Interactive Algorithm for Grammar Inference. In: Proceedings of the Thirteenth National Conference on Artificial Intelligence. AAAI Press. Vol. 2. Pp. 1397.

47. Parekh, R., Yang, J. & Honavar, V. (1996). In: Proceedings of the Thirteenth National Conference on Artificial Intelligence. AAAI Press. Vol. 2. Pp. 1398.

48. Honavar, V. (1993). Learning with Symbolic and Subsymbolic Representations: Some Possibilities for Vision. In: Proceedings of the AAAI Fall Symposium on Machine Learning in Computer Vision. Raleigh, North Carolina. (Also published as AAAI Tech. Rep. FS 93-04). Pp. 162-166.
49. Honavar, V. (1992). Generalized Distance Measures – A Basis for the Integration of Symbolic and Connectionist Learning. In: Workshop on Integrating Neural and Symbolic Processes – The Cognitive Dimension. AAAI-92, San Jose, California.
50. Honavar, V. (1992). Symbolic and Sub-symbolic Computation in Biological Neural Circuits and Systems. In: Neural Information Processing Systems Post-Conference Workshop on Symbolic and Sub-symbolic Computation in Biological Neural Circuits and Systems. Vail, Colorado.
51. Honavar, V. (1991). Generative Learning in Generalized Connectionist Networks. In: Constructive Induction Session – Eighth International Workshop on Machine Learning. Evanston, IL.
52. Honavar, V. (1991). Language and Knowledge: Communication, Acquisition, and Evolution. Invited presentation in: Second International Workshop on Human and Machine Cognition. Perdido Key, Florida.
53. Mikler, A., Honavar, V. & Wong, J. (1992). A Knowledge-Based Approach to Dealing With Uncertain and Incomplete Information in Communication Network Management. In: Proceedings of the First Canadian Workshop on Uncertainty Management: Theory and Practice. Vancouver, B. C., Canada. Pp. 30-38.
54. Honavar, V. (1991). Toward Integrated Models of Natural Language Evolution, Development, Acquisition, and Communication in Multi-Agent Environments. In: Powers, D. and Reeker, L. (Ed.) Proceedings of the AAAI Spring Symposium on Machine Learning of Natural Language and Ontogeny. (MLNLO '91) pp. 82-86. Kaiserslautern, Germany: German AI Centre(DFKI).
55. Honavar, V. (1990). Toward Generalized Connectionist Networks: An Integration of Symbolic and Sub-Symbolic Approaches to the Design of Intelligent Systems. In: AAAI-90 Workshop on the Integration of Symbolic and Neural Processes. Boston, MA.
56. Honavar, V. (1990). Generative Learning Algorithms for Connectionist Networks. In: NIPS-90 Post-Conference Workshop on Constructive and Destructive Learning Algorithms. Keystone, CO.

Invited Book Reviews

1. Honavar, V. (1990). Parallel Distributed Processing: Implications for Psychology and Neurobiology. Invited review. Connection Science.

2. Honavar, V. (1992). Neural Network Design and the Complexity of Learning. Invited review. Machine Learning 9 95-98.

Theses and Dissertations

1. Honavar, V. (1990). Generative Learning Structures and Processes for Generalized Connectionist Networks. Doctoral Dissertation. Madison, WI: Computer Sciences Dept. University of Wisconsin-Madison. Advisor: Professor Leonard Uhr.
2. Honavar, V. (1984). Automated Analysis of Dark-Field Autoradiographs. Masters Thesis. Philadelphia, PA: Center for Image Processing and Pattern Recognition. Department of Electrical and Computer Engineering. Drexel University. Advisor: Professor Oleh Tretiak.

Selected Invited Colloquia

1. **Keynote Talk,** On the importance of causal models in making sense of big data, US-Serbia and West Balkan Workshop on Data Science, Belgrade, Serbia. 2018
2. **Plenary Talk,** Data Science Meets Brain Science, Workshop on Brain, Computation, and Learning, Indian Institute of Science, 2018.
3. **Plenary Talk,** Transdisciplinary Foundations of Data Science. Workshop on Brain, Computation, and Learning, Indian Institute of Science, 2017.
4. **Plenary Talk,** Data Analytics and the Internet of Things. U of Albany President's Forum on Data. 2017.
5. **Plenary Talk,** Transdisciplinary Foundations of Data Science: Mind the Gaps. Institute for Mathematics and Its applications, University of Minnesota, 2016.
6. **Invited Lecture**, From Data Analytics to Computational Discovery and Discovery Informatics, Arizona State University. 2013.
7. **Bortree Lecture,** Computational Prediction of Protein Interfaces and Interactions, Pennsylvania State University. 2013.
8. **Distinguished Lecture**, Computational Prediction of Protein Interfaces and Interactions, Georgia State University. January 2013.
9. **Distinguished Lecture**, From Big Data Analytics to Discovery Informatics. Pennsylvania State University, December 2012.
10. **Invited Colloquium,** Computational Prediction of Protein Interfaces and Interactions, University of California Irvine, September 2012.
11. **Invited Colloquium,** Computational Prediction of Protein Interfaces and Interactions. University of North Texas, June 2012.
12. **Invited Talk,** Towards Infrastructure for Collaborative Discovery. IciS Workshop on Integrating, Representing, and Reasoning over Human Knowledge, August 2010.
13. **Invited Talk**, Knowledge Acquisition from Semantically Disparate, Distributed Data. CISE (IIS), National Science Foundation, May 2010.
14. **Invited Lecture**, From Annotating Sequences to Aligning Networks. Computation in Biology and Medicine Annual Retreat, University of Wisconsin-Madison, October 2009.
15. **Invited Colloquium,** Transforming Biology From a Descriptive Science into a Predictive Science, Indian Institute of Information Technology, Bangalore, India, January 2009.
16. **Invited Colloquium,** Transforming Biology From a Descriptive Science into a Predictive Science: Predictive Models of Macromolecular Function and Interaction. Bioinformatics Center, University of Pune, India, December 2008.
17. **Invited Colloquium,** Semantics-Enabled Infrastructure for Collaborative, Integrative e- Science. School of Information Technology, Jawaharlal Nehru University, New Delhi, India, December 2008.
18. **Invited Talk,** Computational Sciences. High Performance Computing Center, Jawaharlal Nehru University, New Delhi, India, December 2008.

19. **Invited Colloquium,** Semantics-enabled infrastructure for collaborative, integrative e-science. Yahoo!, Bangalore, India, January 2008.
20. **Invited Colloquium,** Algorithms and Software for Knowledge Acquisition from Semantically Heterogeneous, Distributed Data Sources. Dept. of Electrical and Computer Engineering. University of Iowa. 2006.
21. **Invited Colloquium,** Algorithms and Software for Collaborative Discovery in Systems Biology. Dept. Biostatistics, Bioinformatics & Epidemiology. Medical University of South Carolina, 2006.
22. **Invited Talk,** Algorithms and Software for Knowledge Acquisition from Semantically Heterogeneous, Distributed, Autonomous Information Sources. Google Research, 2005.
23. **Invited Talk,** All Science is Computer Science. Iowa Undergraduate Consortium. Drake University, 2004.
24. **Invited Colloquium,** Computational Discovery of Protein Sequence-Structure-Function Relationships: Bioinformatics Infrastructure and Sample Applications. University of Wisconsin-Madison Biostatics and Medical Informatics Department. 2002.
25. **Invited Talk,** Algorithmic and Systems Approaches to Computer Assisted Knowledge Discovery from Biological Data. Iowa State University – University of Iowa Joint Workshop on Bioinformatics. November 3-4, 2000.
26. **Invited Talk,** Neuromimetic Adaptive Autonomous Intelligent Systems. Institute for Computer Applications in Science and Engineering. NASA-Langley Research Center. Hampton, VA. September 28, 1999.
27. **Invited** Colloquium, Kolmogorov Complexity and Computational Learning Theory: Some Emerging Connections and Recent Results. Center for Neural Basis of Cognition*,* Carnegie Mellon University and University of Pittsburgh, Pittsburgh, PA. 1998.
28. **Invited Colloquium,** Experiments in Evolutionary Robotics. Department of Mathematics and Computer Science, Grinnell College*,* Iowa. October 1996.
29. **Invited Lecture,** Data Mining and Knowledge Discovery. Irish Life*,* Des Moines, Iowa. September 1996.
30. **Invited Lecture,** Knowledge Acquisition through Machine Learning. Principal Mutual*,* Des Moines, Iowa. January 1994.
31. **Invited Colloquium,** Generalized Connectionist Networks and Processes for Intelligent Systems. International Computer Science Institute, Berkeley, *CA.* (1990).
32. **Invited Colloquium,** Generative Learning Structures and Processes for Generalized Connectionist Networks. Cognitive and Learning Systems Laboratory, Siemens Research*,* Princeton, NJ. (1990).

Invited Tutorials

1. Honavar, V. Tutorial: Semantics-Enabled e-Science Cyberinfrastructure for Data Mining, 15th Italian Symposium on Advanced Database Systems. Torre Canne, Italy.
2. Honavar, V. Tutorial: Machine Learning Approaches in Computational and Systems Biology. International Conference on Intelligent Sensing and Information Processing, Bangalore, India, 2006

3. Honavar, V. and Caragea, D. Tutorial: Semantic Web for Collaborative Knowledge Acquisition, IEEE International Conference on Digital Information Management, Bangalore, India, 2006.
4. Honavar, V. and Caragea, D. Tutorial: Collaborative Knowledge Acquisition from Semantically Disparate, Distributed Data Sources, 2006 International Symposium on Collaborative Technologies and Systems, Las Vegas, Nevada, USA, May 2006.
5. Honavar, V. and Caragea, D. Semantic Web Technologies for Collaborative Knowledge Acquisition, International Conference on Digital Information Management, Bangalore, India, December 2006.
6. Honavar, V. Intelligent Agents and Multi-Agent Systems IEEE Conference on Evolutionary Computation (CEC), Washington, DC. 1999.
7. Honavar, V. Computational Learning Theory, Genetic Programming Conference, Stanford, 1997.
8. Honavar, V. Intelligent Agents, Genetic Programming Conference, Madison, WI, 1998.

Invited Panel Presentations

1. Panel on Data Science Training for Social and Behavioral Sciences, National Institutes of Health, 2017.
2. Panel on Career Pathways for PhDs in AI and Related Areas. SIGAI Career Network and Conference, Boston, MA 2016.
3. Panel on Big Data, Dupont Summit, Washington DC, 2013.
4. Panel on Big Data Research Opportunities, IEEE International Conference on Big Data, 2013.
5. Panel on Big Data and Discovery Informatics, AAAI Fall Symposium on Discovery Informatics, Washington DC, November 2012.
6. Panel on Big Data Research Opportunities and Challenges in Biological Sciences, ACM Conference on Bioinformatics and Computational Biology, Orlando, Florida, 2012.
7. Panel on Digital Humanities and Computer Science, Chicago Colloquium on Digital Humanities and Computer Science, Chicago, November 2009.
8. Panel on Data Mining Careers, SIAM Conference on Data Mining (SDM 2009), April 2009.
9. Panel on Interdisciplinary Research and Training in Bioinformatics and Computational Biology, Annual Conference of the Italian Association for Artificial Intelligence (AI*IA 2008), Cagliari, Italy, September 2008.
10. Panel on Semantic Data Integration. NSF Workshop on Biomedical Informatics, 2007.
11. Panel on Learning in Knowledge-Based Systems. Second World Congress on Expert Systems. Lisbon, Portugal (1994).
12. Panel on Hybrid Architectures for Intelligent Systems. Second World Congress on Expert Systems. Lisbon, Portugal (1994).
13. Panel on Hybrid Intelligent Systems (SIGHI meeting) World Congress on Neural Networks. San Diego, U.S.A. (1994).

SOFTWARE, DATABASES, SERVERS

- INDUS: Algorithms and Software for Learning Predictive Models from Distributed Data:
  - Indus Learning Framework http://code.google.com/p/induslearningframework/

    - Indus Integration Framework http://code.google.com/p/indusintegrationframework/
- iPref-R: A Qualitative Preference Reasoner: http://fmg.cs.iastate.edu/project-pages/preference-reasoner/
- Biomolecular Network Alignment Toolkit: http://ailab1.ist.psu.edu/BinaWebApp/
- Bionetworkbench: http://bionetworkbench.sourceforge.net/
- Database of Protein-protein Interfaces: http://ailab1.ist.psu.edu/protInDb/index.py
- Database of Protein-RNA Interfaces: http://pridb.gdcb.edu
- Gennotate: A Genome Annotation Toolkit. http://ailab.ist.psu.edu/gennotate/
- NPS-HomPPI: Non Partner-Specific Sequence Homology-Based Protein-Protein Interface Prediction Server: http://ailab1.ist.psu.edu/NPSHOMPPI/
- PS-HomPPI: Partner-Specific Sequence Homology Based Protein-Protein Interface Prediction Server: http://ailab1.ist.psu.edu/PSHOMPPIv1.3/
- PrISE: A Local Structural Similarity Based Protein-Protein Interface Prediction Server: http://ailab1.ist.psu.edu/prise/index.py
- DockRank: Ranking Docked Models Based on Predicted Interfaces: http://ailab1.ist.psu.edu/DockRank/
- ANExDB: Animal Gene Expression Database: http://www.anexdb.org/
- Protein-RNA Interface Prediction Server: http://ailab1.ist.psu.edu/FastRNABindR/
- MHC-II Binding Affinity Prediction Server: http://ailab.ist.psu.edu/mhcmir/
- B-cell epitope Prediction Server: http://ailab.ist.psu.edu/bcpred/
- Epitope Prediction Toolkit: http://ailab.ist.psu.edu/epit/

STUDENT MENTORING

I find working with graduate and undergraduate students to be an extremely rewarding experience. I work with exceptional Ph.D. students with diverse backgrounds – ranging from very theoretical to very experimental. I also enjoy working with M.S. students and undergraduates interested in research. My students benefit from strong mentoring and close interactions within a very collaborative research group. Research-based training in my research group emphasizes development of skills and expertise necessary for the pursuit of a successful independent research career:

(a) Ability to identify and formulate fundamental research problems;

(b) Ability to critically review scientific work;

(c) Ability to conceive, plan, and propose research projects;

(d) Ability to develop creative and innovative solutions;

(e) Strong experimental and/or theoretical expertise in relevant areas;

(f) Effective writing and presentation skills;

(g) Ability to develop and sustain productive research collaborations;

(h) Strong sense of ethics and responsibility in conduct of research.

Fundamental scientific questions (e.g., what is the algorithmic basis of cumulative multi-task learning? How is information encoded, stored, retrieved, decoded, and used in biological systems? How can we precisely characterize the syntax and semantics of the language of macromolecular sequences?); or important practical problems (how do we extract, assimilate, and use information from heterogeneous, distributed, autonomous data and knowledge sources to facilitate collaborative scientific discovery in biology?) drive our research.

My group takes a problem-centered approach to research. In addition to all the usual requirements for successful research, this requires a willingness to acquire, adapt, develop, and apply techniques and tools from areas that lie outside the traditional boundaries of the discipline (e.g., Computer Science) or a sub-discipline (e.g., Machine Learning) when necessary to solve a research problem.
Graduate students who join my lab typically have a broad-based training in Computer Science or a closely related discipline. Many have a strong interest in developing algorithmic or computational models of intelligent behavior (including learning and multi-agent interaction). Some have an interest in developing and applying algorithmic tools for scientific discovery in computational biology and bioinformatics. Some have an interest in building scalable, flexible, extensible, robust, and open-ended distributed information systems. I encourage and nurture interaction among members of my group through research seminars and collaborative research projects.

All of my former Ph.D. students have taken up academic careers or research-oriented careers in the industry. M.S. graduates typically end up in industry. Undergraduates who have worked in my lab often pursue graduate study at one of the other universities with strong programs in Artificial Intelligence or a related area (e.g., Computational Biology).

GRADUATE STUDENT AND POSTDOCTORAL RESEARCH SUPERVISION

Postdoctoral Fellows (Past)

1. **Donna Coffman** (2015-2016). Supported by an NIH K-award. Current position: Assistant Professor. Temple University.
2. **Ganesh Ram Santhanam** (2011-2013), Research Associate, Center for Computational Intelligence, Learning, and Discovery, Iowa State University. Topics: Representing and reasoning about preferences. Model checking approaches to finding and reasoning about intervention policies in networks. Current Position: Postdoctoral Research Associate, Computer Science, Iowa State University.
3. **Jia Tao** (2013) Research Associate, Center for Computational Intelligence, Learning, and Discovery. Iowa State University. Topics: Epistemic Description Logics, Secrecy-preserving Query Answering. Current Position: Visiting Assistant Professor, Bryn Mawr College, Pennsylvania.
4. **Yasser El-Manzalawy (2010-2011).** Research Associate, Center for Computational Intelligence, Learning, and Discovery (CCILD), Iowa State University and Center for Big Data Analytics and Discovery Informatics (Pennsylvania State University). Topic: Machine learning approaches in Immune Informatics. And College of Information Sciences and Technology, Pennsylvania State University. Topic: Machine Learning and Biomedical Big Data Analytics. Yasser is supported in part by a Clinical and Translational Sciences Institute award from the National Institutes of Health.
5. **Li Xue** (2012-2013) Research Associate, Center for Computational Intelligence, Learning, and Discovery (CCILD), Iowa State University. Topic: Computational Prediction of Protein-Protein and Protein-RNA Interfaces and Interactions. Li was supported by CCILD.
6. **Cornelia Caragea** (2010) Research Associate, Center for Computational Intelligence, Learning, and Discovery (CCILD), Iowa State University. Topic: Machine Learning in Computational Biology.

Cornelia was supported in part on a grant from the NSF and in part by CCILD. Current position: Research Associate, Pennsylvania State University.

7. **George Voutsadakis,** Research Associate, Center for Computational Intelligence, Learning, and Discovery (CCILD), Iowa State University. Topics: Federated Knowledge Bases, Semantic Web, Secrecy-Preserving Reasoning. George was supported in part by the ISU NSF Industry-University Cooperative Research Center in Computer and Network Security and in part by CCILD. Current position: Associate Professor of Mathematics and Computer Science, Lake Superior State University.
8. **Jae-Hyung Lee** (2008-2009). Research Associate, Center for Computational Intelligence, Learning, and Discovery (CCILD), Iowa State University. Topic: Construction and Analysis of Macromolecular Interaction Networks. Jae-Hyung was supported in part by CCILD. Current Position: Research Associate, UCLA.
9. **Jie Bao** (2007-2008). Research Associate, Center for Computational Intelligence, Learning, and Discovery (CCILD), Iowa State University. Topic: Representing and Reasoning with Federated Ontologies: Selective Knowledge Reuse, Privacy-preserving reasoning. Jie was supported in part by a grant from the National Science Foundation and in part by CCILD. Current Position: Research Associate (with Jim Hendler), RPI.
10. **Doina Caragea** (2004-2006) Research Associate, Center for Computational Intelligence, Learning, and Discovery (CCILD), Iowa State University. Topic: Knowledge Acquisition from Semantically Heterogeneous, Distributed Information Sources. 2004-2006. Doina is supported in part by a grant from the National Institutes of Health and in part by CCILD. Current Position: Assistant Professor of Computer Science, Kansas State University.
11. **Byron Olson** (2005-2006). Research Associate, Center for Computational Intelligence, Learning, and Discovery (CCILD), Iowa State University. Topic: Discovery of Macromolecular Sequence-Structure-Function Relationships, Transcriptomics, Proteomics, and Interactomics. Byron was supported in part by a grant from the National Institutes of Health and in part by CCILD. Current Position: Research Assistant Professor of Electrical and Computer Engineering, Arizona State University.

Post-Doctoral Fellows (Current)

1. **Justin Petucci.** Research Associate, Institute for Computational and Data Sciences. Topic: Predictive and causal modeling of health risks and health outcomes from electronic health records and other data sources.

Ph.D. Graduates

1. **Sam Gur** (Neuroscience, Pennsylvania State University); 2019. Machine-learning-based functional connectivity analyses. Graduation: 2019. Sam was funded in part by a teaching assistantship and in part by a research assistantship supported by the Edward Frymoyer Endowed Chair. Current Employment: Startup.
2. **Sangchack Lee** (Information Sciences and Technology, Pennsylvania State University); 2018. Causal Inference from Relational Data. Sanghack is supported in part by a research assistantship funded by the National Science Foundation. Current Employment: Columbia University.
3. **Ngot Bui** (Information Sciences and Technology, Pennsylvania State University); 2016; Labeling Actors and Uncovering Causal Accounts of their States in Social Networks and Social Media. Ngot was supported in part by a research assistantship funded by the National Science Foundation. Current Employment: Google.

4. **Rasna Walia** (Bioinformaytics and Computational Biology; Iowa State University; With Drena Dobbs); 2014. Sequence-based prediction of RNA-protein interactions. Carson was supported in part by a research assistantship funded by the National Science Foundationand a Teaching Assistantship in Computer Science. First Employment: Research Scientist, USDA, Ames,Iowa. Current employment: Data Scientist, Corteva Agriscience.
5. **Carson Andorf** (Bioinformatics & Computational Biology; Iowa State University; with Drena Dobbs), 2013. Discovering meaning from biological sequences: focus on predicting misannotated proteins, binding patterns, and G4-quadruplex secondary structures. Carson was supported by an IGERT fellowship funded by the National Science Foundation and a research assistantship funded by the National Institutes of Health. Current employment: Senior Research Scientist, USDA, Ames, Iowa.
6. **Harris Lin** (Computer Science, Iowa State University). Learning Classifiers from Linked Data. Harris was supported in part by a research assistantship funded in part by a grant from the National Science Foundation. Current Employment: Research Scientist, US DOE Ames Laboratory, Ames, Iowa.
7. **Li Xue** (Bioinformatics and Computational Biology, Iowa State University. With Drena Dobbs). Sequence-Homology Based Methods for Protein Interface Prediction and their Application to Ranking Docked Conformations. Graduate Research Excellence Award, Ph.D., 2012. Li was supported in part by research assistantships funded by a grant from the National Institutes of Health and in part by the Iowa State University Center for Computational Intelligence, Learning, and Discovery. Initial Employment: Postdoctoral Research Associated, Iowa State University. Current Employment: Assistant Professor of Bioinformatics, Radboud University.
8. **Hsien-chao Chou** (Bioinformatics & Computational Biology; with Volker Brendel), Local assembly and pre-mRNA splicing analysis by high-throughput sequencing data. Ph.D., 2012. Current Employment: Senior Bioinformatics Scientist, National Cancer Institute, NIH.
9. **Raphael Jordan** (Computer Science, Iowa State University), Ph.D., 2012. Structure-Based Prediction of Protein-Protein Interaction Sites. Graduate Research Excellence Award. Raphael was supported in part by a research assistantship funded by a grant from the National Science Foundation and in part by a teaching assistantship in Computer Science. Current Employment: Data Mining Research Scientist, Quantcast.
10. **Kewei Tu** (Computer Science, Iowa State University), Ph.D., 2012. Unsupervised Learning of Probabilistic Grammars. Kewei was supported by in part by a teaching assistantship in Computer Science and in part by a research assistantship funded by a grant from the National Science Foundation. Initial Employment: Postdoctoral Research Associate, Department of Statistics and Department of Computer Science, University of California Los Angeles. Current Employment: Associate Professsor of Computer Science, Shanghai Tech.
11. **Jia Tao** (Computer Science, Iowa State University; with Giora Slutzki). Topics in Knowledge Bases: Epistemic Description Logics and Secrecy-preserving Reasoning. Ph.D., 2012. Initial Employment, Postdoctoral Research Associate, Iowa State University. Current Employment: Assistant Professor of Computer Science, Lafayette College.
12. **Fadi Towfic** (Bioinformatics and Computational Biology, Iowa State University; with **M. Heather West Greenlee**). Ph.D., 2011. Modular Algorithms for Biomolecular Network Alignment. Fadi was supported in part by a fellowship funded by an Integrative Graduate Education and Research Training (IGERT) grant from the National Science Foundation. Initial employment: Postdoctoral Research Associate, The Broad Institute of MIT and Harvard. Current Employment: Director of Bioinformatics, Bristol Myers Squibb.
13. **Neeraj Koul** (Computer Science, Iowa State University). Ph.D., 2011. Learning Predictive Models from

Massive, Semantically Disparate Data. Neeraj was supported in part by a research assistantship funded by a grant from the National Science Foundation. Initial Employment: Machine Learning Research Scientist, GILT Group. Current Employment: Senior Principal Scientist, AI at MITRE.

14. **Ganesh Ram Santhanam** (Computer Science, Iowa State University), Ph.D., 2010. Representing and Reasoning with Qualitative Preferences for Compositional Systems. Ganesh was supported in part by research assistantships funded by a grant from the National Science Foundation. Initial Employment: Postdoctoral Research Associate, Center for Computational Intelligence, Learning, and Discovery, Iowa State University. Current Position: Associate Scientist, Iowa State University.

15. **George Voutsadakis** (Computer Science, Iowa State University; with Giora Slutzki), Ph.D., 2010. Federated Description Logics for the Semantic Web. George was supported in part by research assistantships funded by a grant from the National Science Foundation. Current Position: Associate Professor of Mathematics and Computer Science, Lake Superior State University.

16. **Cornelia Caragea** (Computer Science, Iowa State University), Ph.D., 2009. Abstraction-Based Probabilistic Models for Sequence Classification. Cornelia was supported in part by research assistantships funded by grants from the National Science Foundation and the National Institutes of Health. Initial Position: Postdoctoral Research Associate, Center for Computational Intelligence, Learning, and Discovery, Iowa State University. Current Position: Associate Professor of Computer Science, University of Illinois at Chicago.

17. **Oksana Yakhnenko** (Computer Science, Iowa State University), Ph.D., 2009. Learning from Text and Images: Probabilistic Generative and Discriminative Models for Partially Labeled Data. Oksana was supported in part by a graduate teaching assistantship in Computer Science and in part by a research assistantship funded by a grant from the National Science Foundation. Initial Position: Postdoctoral Research Associate, INRIA, Rhone-Alpes, France. Current Position: Software Engineer, Google.

18. **Kent Vander Velden** (Bioinformatics & Computational Biology, Iowa State University; with Peter Reilly), Ph.D., 2009. Modeling, Simulation, Synthesis, and Optimization of Biochemical Networks. Kent was supported in part by a graduate fellowship funded by an NSF IGERT award and by Pioneer Hi-Bred. Initial Position: Research Scientist, Pioneer Hi-Bred. Current Position: CEO, Confluence R&D.

19. **Yasser El-Manzalawy** (Computer Science, Iowa State University), Ph.D., 2008. Thesis: Machine Learning Approaches to Epitope Prediction. Recipient of an Iowa State University Graduate Research Excellence Award. Yasser was supported by a fellowship from the Egyptian Government. Current Position: Assistant Professor, Geisinger.

20. **Michael Terribilini**, (Bioinformatics & Computational Biology, Iowa State University; with Drena Dobbs), Ph.D., 2008. Thesis: Deciphering the Protein-RNA Recognition Code. Michael was supported by a Multidisciplinary Graduate Education and Training (MGET) fellowship funded by the USDA. Current Position: Associate Professor of Biology, Elon College, North Carolina.

21. **Feihong Wu** (Bioinformatics & Computational Biology, Iowa State University; with Robert Jernigan), Ph.D., 2009. Thesis: Protein-Protein Interfaces: Databases, Analysis, and Prediction. Feihong was supported by research assistantships funded by the Iowa State University Graduate College and a grant from the National Institutes of Health. Initial Position: Research Scientist, Yahoo! Inc. Current Position: Software Engineer, Salesforce.

22. **Adrian Silvescu** (Computer Science, Iowa State University), Ph.D., 2008; Thesis: Inductive Learning Via Abstraction and Superstructuring. Adrian was supported through a teaching assistantship from the Department of Computer Science and research assistantships funded by Pioneer Hi-Bred and the National Science Foundation. Initial Position: Research Scientist, Yahoo! Inc. Current Position: Senior Research Scientist, Naviance.

23. **LaRon Hughes** (Bioinformatics and Computational Biology, Iowa State University; with James Reecy). Ph.D., 2008. Thesis: Design, Implementation, and Some Applications of an Animal Trait Ontology. LaRon was supported in part by an NSF Integrative Graduate Education and Research Training (IGERT) Fellowship in Bioinformatics and Computational Biology. Initial Employment: Research Scientist, Genome Quest.

24. **Jie Bao** (Computer Science, Iowa State University). Ph.D., 2007. Thesis: Representing and Reasoning with Modular Ontologies. Jie Bao was supported by a research assistantship in Computer Science funded in part by the National Science Foundation, the Iowa State University Center for Integrative Animal Genomics, and the Center for Computational Intelligence, Learning, and Discovery. Initial Position: Research Associate, Computer Science, Rensselaer Polytechnic Institute. Current Position: CEO, Memect.

25. **Tyra Dunn** (Bioinformatics and Computational Biology, Iowa State University; with M. Heather West Greenlee). Ph.D., 2007. Thesis: Characterizing and Influencing Differentiation of Retinal Progenitor Cells. Tyra was supported in part by an Integrative Graduate Education and Research Training (IGERT) fellowship from the National Science Foundation.

26. **Jyotishman Pathak** (Computer Science, Iowa State University). Ph.D., 2007. Thesis: Interactive and Verifiable Web Service Composition, Reformulation, and Adaptation. Recipient of an Iowa State University Graduate Research Excellence Award. Jyotish was supported by a graduate research assistantship in Computer Science funded in part by grants from the National Science Foundation. Initial Employment: Research Scientist, Division of Biomedical Informatics, Mayo College of Medicine, Rochester, Minnesota. Current Employment: Chief of the Division of Health Informatics, Weill-Cornell College of Medicine, Cornell University.

27. **Dae-Ki Kang** (Computer Science, Iowa State University). Ph.D., 2007. Abstraction, Aggregation, and Refinement Strategies for Machine Learning. Dae-Ki was funded by a Teaching assistantship from the Department of Computer Science and a Research assistantship funded by the National Science Foundation. Current Employment: Associate Professor, Department of Computer Engineering, Dongseo University, Pusan, Korea.

28. **Jun Zhang** (Computer Science). Ph.D., 2005. Ontology Aware Learning Algorithms. Recipient of an Iowa State University Graduate Research Excellence Award. Jun was supported in part by research assistantship funded by a grant from the National Science Foundation and a teaching assistantship from the ISU Department of Computer Science. Initial Employment: Research Scientist, Fair Isaac, San Diego. Current Employment: Chief Data Scientist at ThreatMetrix

29. **Changhui Yan** (Bioinformatics and Computational Biology, Iowa State University; with Drena Dobbs). Ph.D., 2004. Computational Approaches to Prediction of Protein-Protein, Protein-DNA and Protein-RNA Interfaces. 2005. Recipient of an Iowa State University Graduate Research Excellence Award. Changhui was supported by a Plant Sciences Fellowship and research assistantships funded by the ISU Graduate College and a grant from the National Institutes of Health. Initial Employment: Assistant Professor of Computer Science, Utah State University. Current position: Professor of Computer Science, North Dakota State University.

30. **Doina Caragea** (Computer Science, Iowa State University). Ph.D., 2004. Learning Classifiers from Semantically Heterogeneous, Distributed, Autonomous Data Sources. 2004. Recipient of an Iowa State University Graduate Research Excellence Award. Initial Employment: Research Associate, Center for Computational Intelligence, Learning, and Discovery, Iowa State University. Current position: Professor of Computer Science, Kansas State University.

31. **Jihoon Yang** (Computer Science, Iowa State University). Ph.D., 1999. Learning Agents for Information Retrieval and Knowledge Discovery, 1999. Initial Employment: Research Scientist, Information Sciences Laboratory, Hughes Research Laboratory, Malibu, CA. Current Employment: Professor of Computer Science, Sogang National University, Korea.
32. **Karthik Balakrishnan** (Computer Science, Iowa State University). Ph.D., 1998. Biologically Inspired Computational Structures and Processes for Autonomous Agents and Robots, Recipient of an Iowa State University Graduate Research Excellence Award. Initial Employment: Senior Research Scientist, Datamining Group, Allstate Research and Planning Center. Current Position: Senior Vice President of IoT, Verisk.
33. **Rajesh Parekh** (Computer Science, Iowa State University). Ph.D., 1998. Recipient of an Iowa State University Graduate Research Excellence Award. Machine Learning of Automata and Neural Network Pattern Classifiers, 1998. Initial Employment: Senior Research Scientist, Data mining Group, Allstate Research and Planning Center, Menlo Park, CA. USA. Current position: Engineering Director, Google.
34. **Chun-Hsien Chen** (Computer Science, Iowa State University). Ph.D., 1997. Neural Architectures for Knowledge Representation and Inference, 1997. Initial Employment: Research Scientist, Advanced Technology Center, Computer and Communication Research Laboratories, Industrial Technology Research Institute, Taiwan. Current Employment: Associate Professor, Department of Information Management, Chang Gung University, Taiwan.
35. **Armin Mikler** (Computer Science, Iowa State University; with Johnny Wong), Ph.D., 1995. Quo Vadis – A Framework for Intelligent Routing in Large Communication Networks. Initial Employment: Research Associate, Scalable Computing Laboratory, DOE Ames Lab, Ames, Iowa. Current Position: Professor of Computer Science, Georgia State University, USA.

Current Ph.D. Students

1. **David Foley,** Informatics, Pennsylvania State University. Interests: Biomedical Data Sciences, Longitudinal Data Analysis. Expected Graduation: Spring 2022. David is supported in part by an NIH-funded BD2K Predoctoral Fellowship in Biomedical Data Sciences.
2. **Junjie Liang,** Informatics, Pennsylvania State University. Interests: Machine Learning. Big Data Analytics. Longitudinal Data Analysis. Expected Graduation: Spring 2022. Junjie is funded in part by a research assistantship funded by the Center for Big Data Analytics and Discovery Informatics.
3. **Yong Jung,** Bioinformatics and Genomics, Pennsylvania State University. Interests: Bioinformatics. Expected graduation: Spring 2018. Yong is supported in part by a research assistantship funded by the National Science Foundation.
4. **Aria Khademi,** Information Sciences and Technology, Pennsylvania State University. Interests: Machine Learning, Big Data Analytics. Expected graduation: Spring 2020. Aria is supported by a research assistantship funded by the Edward Frymoyer Endowed Professorship (held by Vasant Honavar) and in part by a teaching assistantship.
5. **Thanh Le**, Informatics, Pennsylvania State University. Interests: Machine Learning, Big Data Analytics, Biomedical Informatics. Expected graduation: Spring 2020. Thanh is supported by a research assistantship funded by the Center for Big Data Analytics and Discovery Informatics.
6. **Tsung-Yu Hsieh**, Computer Science and Engineering, Pennsylvania State University. Interests: Machine Learning. Expected graduation: Spring 2021. Tsung-Yu is supported by a research assistantship funded by the National Science Foundation.
7. **Yiwei Sun,** Computer Science and Engineering, Pennsylvania State University. Interests: Machine Learning, Big Data Analytics. Expected graduation: Spring 2021. Yiwei is supported by a research assistantship funded by the National Science Foundation.
8. **Yalda Fazlalizadeh,** Informatics, Pennsylvania State University. Interests: Machine Learning. Expected graduation: Spring 2024.

Major Professor, M.S.
Current M.S. Students

1. **Rahul Katiki**, Computer Science and Engineering, Pennsylvania State University. In progress.

M.S. Graduates

2. **Cheng-Kai Chen**, (Computer Science and Engineering), Pennsylvania State University. Comparative analyses of networks. M.S. 2020.
3. **Mayank Garg** (Computer Science and Engineering), Pennsylvania State University; Predicting energy efficiency of batteries. with Dan Kifer. M.S. 2017.
4. **Yiming Zhang** (Bioinformatics & Computational Biology, Iowa State University; with Volker Brendel), M.S. 2014.
5. **Sateesh Kodavalli** (Computer Science, Iowa State University). M.S., 2010. Extensible Problem Specific Tutor (xPST): Easy Authoring of Intelligent Tutoring Systems.
6. **Bhavesh Sanghvi** (Computer Science, Iowa State University). M.S., 2010. Identifying and eliminating inconsistencies in mappings across hierarchical ontologies. Current Employment: Microsoft.
7. **Sushain Pandit** (Computer Science, Iowa State University), M.S., 2010. Ontology-guided extraction of structured information from unstructured text: Identifying and capturing complex relationships. Current Employment: IBM.
8. **Tim Alcon** (Bioinformatics & Computational Biology; with Heather Greenlee)**.** Using a Seed- Network to Query Multiple Large-Scale Gene Expression Datasets from the Developing Retina in Order to Identify and Prioritize Experimental Targets. Tim was supported by a Multidisciplinary Graduate Education and Training (MGET) fellowship funded by the USDA.
9. **Lucas Bonansea** (Human-Computer Interaction; with Stephen Gilbert), 2009. 3-d hand gesture recognition.
10. **Flavian Vasile** (Computer Science, Iowa State University), 2008. Uncovering the structure of hypergraphs through tensor decomposition: An application to folksonomy analysis. Flavian was supported in part by a teaching assistantship in Computer Science, Iowa State University. Current Employment: Software Engineer, Yahoo! Inc.
11. **Oksana Kohutyuk** (Computer Science, Iowa State University), 2007. Thesis: Retina Workbench: Database and software tools for comparative analysis and querying of gene and protein networks. Oksana was supported by a research assistantship funded by the National Institutes of Health and a teaching assistantship in Computer Science, Iowa State University. Current position: Software Engineer, Cisco.
12. **Charles Giesler** (Computer Science, Iowa State University), A Java Reinforcement Learning Module for the Recursive Porous Agent Simulation Toolkit: Facilitating study and experimentation with reinforcement learning in social science multi-agent simulations. 2003. Current Position: Lawrence Livermore Labs.
13. **Anna Atramentov** (Computer Science, Iowa State University), A Multi-Relational Decision Tree Learning Algorithm – Implementation and Experiments. 2003. Current position: Ph.D. Student, University of Illinois at Urbana-Champaign

14. **Zhong Gao** (Bioinformatics, Iowa State University), Genome wide recognition of tumor necrosis factor (TNF) like ligands in human and Arabidopsis genomes: A structural genomics approach. 2003 Initial Employment: Post-doctoral fellow, The Center for Cardiovascular Bioinformatics and Modeling, Johns Hopkins University.
15. **Jaime Reinoso-Castillo** (Computer Science, Iowa State University), Ontology-Driven Query-Centric Information Integration from Heterogeneous, Distributed, Autonmous Data Sources for Computer Assisted Scientific Discovery. 2002. Initial Employment: Universidad Javeriana, Colombia.
16. **Hector Leiva** (Computer Science, Iowa State University), Learning Classifiers from Relational Data. 2002. Initial Employment: Research Scientist, Research Scientist, Universidad Nacional de San Luis. Argentina.
17. **Xiaosi Zhang** (Bioinformatics, Iowa State University), Identification of Functionally Related Genes from Gene Expression Data. 2002.Initial Employment: Papajohn Center for Entreprenuership, Ames, Iowa.
18. **Xiangyun Wang** (Bioinformatics, Iowa State University), Data Mining Approach to Discovery of Protein Sequence-Structure-Function Relationships. 2002. Initial Employment: Astra-Zeneca Inc.
19. **Kent Vander Velden** (Bioinformatics, Iowa State University, joint supervision with Gavin Naylor), Spatial Clustering of Differences in Measured Homoplasy with Respect to Protein Structure. 2002. Initial Employment: Pioneer Hi-Bred, Inc.
20. **Neeraj Koul** (Computer Science, Iowa State University), Clustering With Semi-Metrics, 2001. Initial Employment: Motorola.
21. **Dake Wang** (Computer Science, Iowa State University), Data-Driven Generation of Decision Trees for Motif-Based Assignment of Protein Sequences to Functional Families., 2001. Initial Employment: Lumicyte, Inc. Current Employment: Genentech, Inc.
22. **Rushi Bhatt** (Computer Science, Iowa State University), Spatial Learning and Localization: A Computational Model and Behavioral Simulations, 2001. Ph.D. Program, Boston University.
23. **Fajun Chen** (Computer Science, Iowa State University), Learning Information Extraction Patterns from Text, 2000. Initial Employment: Ericsson.
24. **Tarkeshwari Sharma** (Computer Science, Iowa State University), Agent Toolkit for Distributed Knowledge Networks, 2000. Initial Employment: Motorola, Inc.
25. **Asok Tiyyagura** (Computer Science, Iowa State University), Alternative Criteria for Association Rule Mining, 2000. Cisco Systems, Inc.
26. **Di Wang** (Computer Science, Iowa State University), 1997. Mobile Agents for Information Retrieval.
27. **Shane Konsella** (Computer Science, Iowa State University), 1996. Trie Compaction Using Genetic Algorithms. Initial Employment: Hewlett-Packard.
28. **Karthik Balakrishnan** (Computer Science, Iowa State University), 1993. Faster Learning Approximations of Backpropagation by Handling Flat-Spots. Continued as a Ph.D. student.
29. **Jayathi Janakiraman** (Computer Science, Iowa State University), 1993. Adaptive Learning Rate for Increasing Learning Speed in Backpropagation Networks. Initial Empoyment: Motorola.
30. **Priyamvadha Thambu** (Computer Science, Iowa State University), 1993. Automated Knowledge-Base Consistency Maintenance in an Evolving Intelligent Advisory System. Initial Employment: Inference Corporation.

31. **Rajesh Parekh**, 1993 (Computer Science, Iowa State University). Efficient Learning of Regular Languages Using Teacher-Supplied Positive Examples and Learner-Generated Queries. Continued as a Ph.D. student.

32. **Richard Spartz**, 1992 (Computer Science, Iowa State University). Speeding Up Backpropagation Using Expected Source Values. Initial Employment: IBM.

Member of Graduate Program of Study (Thesis) Committees
Ph.D. Committees

1. Chen Liang	Information Sciences and Technology, PSU	2018
2. Yafei Wang	Information Sciences and Technology, PSU	2017
3. Moojan Gafurian	Information Sciences and Technology, PSU	2017
4. Rabah Al-Zaidy	Computer Science and Engineering, PSU	2017
5. Mehdi Bagherzadeh	Computer Science, ISU	2016
6. Seifu Chonde	Industrial Engineering, PSU	2016
7. Robert Dyer	Computer Science, ISU	2013
8. Scott Boyken	Bioinformatics and Computational Biology, ISU	2013
9. Haitao Cheng	Bioinformatics and Computational Biology, ISU	2013
10. Ataur Katebi	Bioinformatics and Computational Biology, ISU	2013
11. Jivko Sinapov	Computer Science, ISU	2013
12. Ru He	Computer Science, ISU	2013
13. Chris Strasburg	Computer Science, ISU	2013
14. Zack Oster	Computer Science, ISU	2013
15. Yao Fu	Bioinformatics and Computational Biology, ISU	2013
16. Oliver Couture	Genetics, ISU	2011
17. Tu-Liang Lin	Computer Science, ISU	2011
18. Brian Patterson	Computer Science, ISU	2011
19. Aimin Yan	Bioinformatics and Computational Biology, ISU	2011
20. Deepak Reyon	Bioinformatics and Computational Biology, ISU	2011
21. Scott Broderick	Material Science and Engineering, ISU	2009
22. Changsung Kang	Computer Science, ISU	2008
23. Jeff Sander	Bioinformatics and Computational Biology, ISU	2008
24. Lei Yang	Bioinformatics and Computational Biology, ISU	2008
25. Facundo Bromberg	Computer Science, ISU	2007
26. Jae-Hyung Lee	Bioinformatics and Computational Biology, ISU	2007
27. Yu Cao	Computer Science, ISU	2007
28. Laura Hecker	Neuroscience, ISU	2007
29. Oleksiy Atramentov	Physics, ISU	2006
30. Xiaonan Li	Industrial and Manufacturing Systems Engineering, ISU	2006
31. Di Wu	Bioinformatics and Computational Biology, ISU	2006
32. Yungok Ihm	Bioinformatics and Computational Biology, ISU	2004
33. Cizhiong Zhang,	Bioinformatics and Computational Biology, ISU	2004
34. Zhong Zhang,	Electrical and Computer Engineering, ISU	2004
35. Haibo Cao	Physics, ISU	2003

36. Brooke Peterson	Genetics	2003
37. Marybeth Gurski	Computer Science	2001
38. Guy Helmer	Computer Science, ISU	2001
39. Robi Polikar	Electrical and Computer Engineering, ISU	2001
40. Vincent Van Acker	Electrical and Computer Engineering, ISU	2000
41. Chun-Fu Chen	Economics, ISU	1999
42. Victoria Bascunana	Chemical Engineering, ISU	1999
43. Guozhong Zhou	Electrical and Computer Engineering, ISU	1998
44. Cheng-Chi Tai	Electrical and Computer Engineering, ISU	1998
45. James Lathrop	Computer Science, ISU	1997
46. Krishna Dhara	Computer Science, ISU	1997
47. Babak Fourouraghi	Computer Science, ISU	1995
48. Timothy Wahls	Computer Science, ISU	1995
49. Chang-Chun Tsai	Industrial and Manufacturing Engineering, ISU	1995
50. Sonmez Rifat	Civil and Constructional Engineering, ISU	1995
51. Richa Agrawala	Computer Science, ISU	1994
52. Bamshad Mobasher	Computer Science, ISU	1994
53. Hun Kang	Electrical and Computer Engg., ISU	1993

M.S. Committees

1. Sandeep Krishnan	Computer Science, ISU	2009
2. Zack Oster	Computer Science, ISU	2009
3. Georgi Batinov	Economics, ISU	2007
4. Inya Nlenanya	Agricultural Engineering, ISU	2005
5. Jing Xu	Psychology, ISU	2005
6. Suxing Cheng	Computer Science, ISU	2005
7. Kyongryun Lee	Computer Science, ISU	2005
8. Haitao Cheng	Computer Science, ISU	2004
9. Patricia Lonosky	Genetics, ISU	2002
10. Mallika Bachan	Statistics, ISU	2002
11. Melinda Vander Velden	Electrical and Computer Engineering, ISU	2002
12. Jeremy Patterson	Computer Science, ISU	2001
13. Sa Lin	Computer Science, ISU	2001
14. Vijay Viswanathan	Electrical and Computer Engineering, ISU	2001
15. Fengmei Liu	Computer Science, ISU	2001
16. Xinhua Dong	Computer Science, ISU	2001
17. Mark Slagell	Computer Science, ISU	2001
18. Hao Dong	Computer Science, ISU	2001
19. Jun Li	Computer Science, ISU	2001
20. Thai-Tin Huang	Computer Science, ISU	2000
21. Ran Liu	Computer Science, ISU	2000
22. Xumei Lu	Computer Science, ISU	2000
23. Nanchang Yang	Computer Science, ISU	2000

24. Peng Han	Botany, ISU	1999
25. Jeffrey Yakey	Computer Science, ISU	1999
26. Sunitha Kothapalli	Electrical Engineering, ISU	2000
27. Raghunandan Havaldar	Computer Science, ISU	1998
28. Guy Helmer	Computer Science, ISU	1998
29. Ngee Jenn Lee	Mechanical Engineering, ISU	1998
30. Laura Nelson	Computer Science, ISU	1998
31. Jibin Xiang	Computer Science, ISU	1998
32. Venkat Naganathan	Computer Science, ISU	1998
33. Dean Stevens	Computer Science, ISU	1998
34. Prashant Pai	Computer Science, ISU	1998
35. Abhinav Rawat	Nuclear Engineering, ISU	1998
36. Rishi Nayar	Computer Science, ISU	1997
37. Marcie Goodman	Computer Science, ISU	1997
38. Jon Schultze-Hewett	Computer Science, ISU	1997
39. Chin Khor	Mechanical Engineering, ISU	1997
40. Qiang-lin Zhao	Computer Science, ISU	1996
41. Chi-Chuan Chen	Agricultural Engineering, ISU	1996
42. Mahesh Subramaniam	Computer Science, ISU	1996
43. Glen Holt	Computer Science, ISU	1996
44. Niranjan Vaidya	Computer Science, ISU	1996
45. Thirumalai Anandapillai	Industrial and Manufacturing Engineering, ISU	1995
46. Thomas DeWulf	Electrical Engineering, ISU	1994
47. Raghav Trivedi	Computer Science, ISU	1994
48. Arun Barboza	Computer Science, ISU	1994
49. Brian Schmidt	Electrical and Computer Engineering, ISU	1993
50. Brian Peterson	Computer Science, ISU	1992
51. Salim Chandani	Industrial and Manufacturing Engineering, ISU	1992
52. Prerana Vaidya	Computer Science, ISU	1991
53. James Wittry	Computer Science, ISU	1991
54. Srinivas Boddu	Electrical Engineering, ISU	1991
55. Simanta Mitra	Computer Science, ISU	1991

UNDERGRADUATE STUDENT RESEARCH SUPERVISION

Supervisor, Undergraduate Honors Project

1. Andres De la Fuente (2018-2019) Topics in Reinforcement Learning. PSU.
2. Benjamin Myers (2016-2017). Topics in Machine Learning. PSU.
3. Preston Soeperanto (2016-2017). Topics in Machine Learning. PSU
4. Dom Mirabile (2014). Topics in Machine Learning. PSU.
5. Oksana Yakhneko (2003-2004). Topics in Machine Learning. ISU
6. Eric Barsness (1993), An Object-Oriented Implementation of a Genetic Algorithms Testbed. ISU
7. Daniel Graves (1992), Parallel Architectures for Artificial Intelligence. ISU

Supervisor, Undergraduate Research

1. Letao Qi, Iowa State University, 2012-2013. Graduate school: Rice University.
2. Lionel Barrow, Bard College, Participant, Summer Institute in Bioinformatics and Computational Systems Biology, ISU 2010
3. Kim Eue, Iowa State University, Summer Research Student, 2010
4. Kiran Gustavson, Iowa State University, Summer Research Student, 2010
5. Ryan Pendergast, College of the Holy Cross, Participant, Summer Institute in Bioinformatics and Computational Systems Biology, ISU, 2010
6. Marjie Volk, Iowa State University, Undergraduate Research Assistant, ISU, 2010-2011
7. Katie Wilkins, Case Western Reserve University, Participant, Summer Institute in Bioinformatics and Computational Systems Biology, ISU 2009
8. Casey Oliver, Penn State University, Participant, Summer Institute in Bioinformatics and Computational Systems Biology, ISU 2009
9. Susan Koons, Texas A&M University, Participant, Summer Institute in Bioinformatics and Computational Systems Biology, ISU 2009.
10. Kevin Yang, Cornell University, Summer Research Student, ISU. 2008.
11. Tyson Williams, Undergraduate Research Assistant, ISU 2007-2008.
12. David Gemperline, Participant, Summer Institute in Bioinformatics and Computational Systems Biology, 2007. ISU
13. Keith Callenberg, Participant, Summer Institute in Bioinformatics and Computational Systems Biology, 2007. ISU. Graduate School: Carnegie Mellon University.
14. Remy Younes, Undergraduate Research Assistant. Topics in Data Integration. 2007. ISU
15. Matt Miller, Topics in Learning Classifiers from Distributed Data, 2006-2007. ISU (Graduate School: MIT)
16. John Leacox, Topics in Data Integration, 2006. ISU
17. Peter Wong, Topics in Collaborative Ontology Development, 2006. ISU
18. Ryan Bruce (2004), Topics in Bioinformatics, ISU
19. Cody Pfau (2003). Topics in Data Mining. ISU
20. Amy Nienaber (2003). ISU. Computational Discovery of Protein-Protein Interactions
21. Matthew Beard (2003). Computational Discovery of Protein-Protein Interactions
22. Diane Schroeder (2001-2002) Data Mining Approaches to Discovery of Protein Sequence Function Relationships, ISU (Graduate School: Stanford University)
23. Kent Vander Velden (1998-1999), Protein Structure Prediction. ISU (Graduate School: Bioinformatics, Iowa State University)
24. Jeremy Ludwig (1996-1997), Topics in Neural Computing.ISU (Graduate School: U. Pittsburgh).

25. David DeYoe (1996-1997), Topics in Cognitive Modelling. ISU
26. Carl Pecinovsky (1996-1997), Constructive Neural Network Learning Algorithms.ISU
27. Brian Walenz, Topics in Genetic Algorithms (Graduate School: Computer Science, University of New Mexico). ISU
28. Gabriel Ki (1996-), Situated Robotics. ISU
29. Todd Lindsey (1995-96), Constructive Neural Network Learning Algorithms. ISU
30. Jouko Rytilahti (1994), Explorations in Evolutionary Algorithms. ISU
31. Leigh McMullen (1993), Adaptive Game-Playing Programs. ISU

Mentor, Freshman Honors Study

1. Marjie Volk (2010) Topics in Machine Learning. ISU
2. Jeffrey Schroeder (1997) Topics in Artificial Intelligence. ISU
3. Matthew Potter (1997) Topics in Artificial Intelligence. ISU
4. Brian George (1994), Topics in Neural Computing. ISU
5. Adam Johnson (1994), Topics in Neural Computing. ISU
6. Marcus Ryan (1993), Topics in Artificial Intelligence. ISU

PRE-COLLEGE STUDENT RESEARCH SUPERVISION

Mentor (for pre-college students)

1. Eric Solan, Nic Dayton, Luke Rolfes, and Julian Sheldahl. Animus Facticius. Adventures in Supercomputing (1998) project. First Place.
2. Sara Karbeling et al., A Computational Model of Animal Spatial Learning Adventures in Supercomputing (1998) project.
3. Sara Karbeling, Kellan Brumback, Anna Keyte, and Angel Sherif (1997), Lateral Inhibition and Sensory Processing in the Limulus Polyphemus Eye*,* Adventures in Supercomputing (AIS-97) project. First place in Iowa, and Second Place in the National Competition.
4. Stephen Lee (1993), Topics in Neural Computing.
5. John Farragher (1992), Topics in Neural Computing.
6. Peter Luka (1991), Topics in Neural Computing.

SPONSORSHIP OF VISITING RESEARCHERS

1. Morteza Jaderyan, Visiting PhD Student, Bu Ali Sina University, Iran. 2016-2017.
2. Professor Jinlong Hu, South China University of Technology, 2016-2017.
3. Professor Jihoon Yang, Sogang University, Korea, 2008.
4. Oswaldo Velesz-Langs, Madrid Polytechnic University, Madrid, Spain. 2003.
5. Professor Yigon Kim (on Sabbatical from Yosu National University, Yosu, Korea 2000-2001) Data Mining and Knowledge Discovery*.*
6. Professor Mok Dong Chung (on Sabbatical from Pukyong National University, Korea, 1999-2000) Agent-based systems and knowledge-based systems*.*
7. Olivier Bousquet (from Ecole Polytechnique, France, 1997) Topics in Cognitive Modeling and Robotics.
8. Codrin Nichitiu (from ENS Lyon, France, Summer 1996), Topics in Machine Learning.
9. Dimitri Kotchetkov (Visitor from Ukraine, Summer 1996), Topics in Robotics.

10. Vadim Kirillov (Fullbright Scholar from Ukraine, 1995), Constraint-based Reasoning under uncertainty.

TEACHING STATEMENT

Teaching Philosophy

My teaching philosophy is perhaps best summed up by a quote from Joseph Chesterton: "The Foundation of teaching is research; and the object of research is teaching, that is, the dissemination of knowledge".

In my view, teaching is an integral and extremely rewarding part of academic life. I like to teach what I like to learn. It is hard to top the joy of grasping a new idea or solving an unsolved problem for the very first time. I subscribe to John Dewey's view of education: "Education is not preparation for life; education is life itself". For me, teaching is not just about communicating what is thought to be known, but cultivating the ability in students to challenge the current state of knowledge, and to venture beyond the current frontiers of knowledge into the unknown, on a life-long journey of learning and discovery. As Richard Dann remarked, "He who dares to teach must never cease to learn". Most of my curriculum development and teaching activities in computer science strongly complement, and are sustained by an active research program.

In my view, teaching computer science is not *just* about teaching students to be competent programmers or creators and users of sophisticated software tools; it is about introducing computer science as a profoundly interesting scientific discipline. Computer science, as a discipline provides us with the best language we have so far – the language of algorithms – precise recipes for describing processes that manipulate information – for modeling biological, cognitive, and social phenomena – just as calculus provided Newton and his successors with a language for modeling physical phenomena.

I was drawn to Computer Science because of the possibility of understanding biological, cognitive, and social phenomena in terms of processes that acquire, store, retrieve, manipulate, and use *information.* I am passionate about introducing students to algorithmic approaches to exploring fundamental questions in biological, cognitive, and social sciences and to the challenges of developing information processing artifacts and software that can dramatically improve our quality of life.

Teaching Style

My personal teaching style involves:

(a) Engaging students so that they become active participants in the learning process rather than passive observers ("What I hear, I forget. What I see, I remember. What I do, I understand" – Confucius).

(b) Setting the stage for the topic of study – We learn what we *almost* already know. Hence I find it useful to introduce a complex idea or a new topic or through a succession of smaller steps, establishing their connection with familiar topics, each building on what my students alreadyknow.

(c) Setting challenging yet realistic goals – In my experience, students respond best, and learn most, from learning goals that are both challenging and achievable. Assignments that are trivial are boring. Assignments that are excessively difficult can be frustrating and intimidating.

(d) Letting students make mistakes and learn from them. Exploration and experimentation are essential to learning and discovery and learning what does not work (and why) is as important as learning what works (and why).

(e) Accommodating different learning styles – Kolb identifies several learning styles. Some learn best from observations and examples; others by acquiring general principles and knowledge of how to

apply them in specific situations; others by deducing and discovering general principles or theories; and yet others by doing – that is, by trying things, making mistakes, and learning from them. I find it useful to develop course materials that exercise multiple learning styles.

Teaching and Curriculum Development Experience

Over the past 30 years, I have designed, developed, and taught undergraduate as well as graduate courses and seminars in artificial intelligence, intelligent agents and multi-agent systems, data sciences, causal inference, machine learning, data mining and knowledge discovery, neural and evolutionary computation, computational learning theory, bioinformatics and computational biology. The material covered in the courses is chosen with an emphasis on concepts that are likely to have a lasting impact on the discipline in the years to come. In addition to introducing students to a core body of knowledge in the areas of study, these courses present such knowledge in the broader context of computer science as an intellectual discipline and to develop the students into creative thinkers and problem-solvers, be it in academic research or advanced technology development.

In addition to developing and teaching courses, I have had substantial experience developing new curricula and programs.

As a founding member and member of the Data Sciences Undergraduate program Task force at Pennsylvania State University, I have co-led the development of an undergraduate program in Data Sciences, a joint initiative of the College of Information Sciences and Technology, School of Electrical Engineering and Computer Science, and the Department of Statistics. The program is organized into three tracks, one focused on Big Data Systems and Algorithms, one on Statistical Foundations, and one on Applied Data Sciences. While the tracks share a set of core courses in the foundations of data sciences, they allow students to specialize in the computational, statistical, or applied aspects of data sciences. Electives allow students to get exposure to an application domain of their choosing (e.g., Social Sciences, Life Sciences, etc.).

I have also led the development of the curriculum for an NIH-funded PhD program in Biomedical Data Sciences at Penn State. The core curriculum includes courses in Machine Learning, Statistics, Scalable Methods, and Data Privacy and Ethics, with advanced electives drawn from graduate courses in Data Sciences from multiple disciplines including Computer Science, Statistics, and Information Sciences and Technology. Biomedical Sciences electives offer exposure to one or more relevant area of life sciences, e.g., genomics, population health, etc.

I have also led the development of the curriculum for a Masters Program in Data Sciences at Penn State. The program is aimed at PhD students in a variety of disciplines who are interested in pursuing a concurrent Masters in Data Sciences to expand their skills in managing and analyzing large and complex data sets in their own disciplines as well as Masters students interested in careers in industry. The program is currently under consideration for implementation at Penn State.

As a founding member of the interdepartmental graduate program in Bioinformatics and Computational Biology (BCB) at Iowa State University, I worked with an interdisciplinary team of biologists to secure an Integrative Graduate Education and Research Training (IGERT) award which helped establish one of the first (and one of the largest, and perhaps one of the strongest Bioinformatics Ph.D. programs in the United States, among universities without medical schools. I have led the development of a set of 4 core courses in Bioinformatics and Computational Biology covering Genome Informatics, Structural Genome Informatics, Functional Genomics, and Systems Biology. I have developed and taught short course modules as part of an NSF-supported Bioinformatics Summer Institute in Bioinformatics and Computational Systems Biology. I have also

contributed to the design of a new undergraduate curriculum in Bioinformatics and Computational Biology, which has been offered at Iowa State University beginning in fall 2007.

The undergraduate and graduate courses that I have developed and taught over the years introduce students to some of the most challenging topics in computer science – involving the application of concepts and tools from the theory of computation, design and analysis of algorithms, and design of software systems in the construction of *intelligent* artifacts: computer programs that represent and reason with and about knowledge, acquire knowledge from interaction with their environment, and discover and use regularities from data.

I have developed and taught a combined upper level undergraduate and introductory graduate course in Artificial Intelligence with emphasis on fundamental problems and approaches in the design of intelligent agents. This course is aimed at introducing students to the foundations of artificial intelligence, including intelligent agents, problem solving, knowledge representation and reasoning (including representing and reasoning with uncertainty, decision theory, planning, and machine learning. This course has been fairly popular with graduate and senior undergraduate students in Computer Science, Engineering, and related disciplines.

I have developed and taught a graduate course in Machine Learning which presents a unified framework for formulation and solution of a broad class of machine learning problems using mathematical tools drawn from probability theory, statistics, information theory, decision theory, and algorithm design and analysis. This course is designed to help students gain a broad understanding of the current state of the art in machine learning, adapt and apply machine learning approaches to real- world applications (e.g., in computational biology, semantic web), and begin to conduct original research in machine learning.

I have developed an undergraduate course on neural computation. Because no textbook on this topic suitable for undergraduate students in Computer Science was available in 1992 when I first offered the course, I developed most of the material for this course in the form of lecture notes some of which were adapted for use in graduate and undergraduate courses in other universities. The focus of the course was on computational models of neurons and networks of neurons (neural networks) and neural network learning algorithms. This course was popular with undergraduates in Computer Science, Engineering, as well as graduate students from other disciplines until the course was eventually superseded by a broad-based introduction to machine learning which I developed and began to teach in 2000.

I have developed and taught a graduate course on intelligent agents and multi-agent systems which draws on material from game theory, knowledge representation and inference, decision theory, contract theory, bargaining theory, and related areas to develop an understanding of fundamental problems in the design of open-ended systems consisting of loosely coupled systems consisting of interacting autonomous entities (information sources, intelligent agents) e.g., the semantic web.

I have also developed and taught advanced graduate "topics in artificial intelligence" courses that have covered computational learning theory, probabilistic graphical models, information retrieval, computational molecular biology, knowledge representation and inference, neural computation, evolutionary algorithms, reinforcement learning, and intelligent agents and multi-agent systems, and semantic web.

I have developed and co-taught a course on Functional Genomics and Computational Systems Biology which is the fourth in a sequence of 4-course core curriculum that I helped develop for the Bioinformatics and Computational Biology graduate program. This course, among the first of its kind, was offered for the first time at ISU in the spring of 2008.

In addition to the regular courses, current research topics are explored in depth in research seminars, which I have organized or co-organized with the help and active participation of my graduate students.

The nature of the material taught in my courses requires a delicate balance between theory and experimentation. In a fast-paced field like computer science in general and artificial intelligence and bioinformatics in particular, the courses have to anticipate key developments in the field that are likely to have a long-term impact and provide students with a solid understanding of the fundamentals as well the insight that comes with hands-on experience. Hence, I have invested in efforts to develop the laboratory facilities that are essential to support experiments, exercises, and projects that enhance the students' understanding the material covered in the courses.

To help develop the written and oral communication skills of students, most of my courses require individual or team research projects culminating in a short paper. It has been my experience that team projects promote collaborative learning and problem solving. The projects often serve as vehicles for integrating latest research results into the graduate and undergraduate curriculum. They also provide an opportunity for students to exercise their creativity and explore new solutions to open problems in artificial intelligence. In many instances, such class projects have evolved into thesis research topics or produced results that were eventually published in refereed national and international conferences.

Teaching Interests

I have a strong interest in teaching undergraduate and graduate courses in artificial intelligence, machine learning, bioinformatics and computational biology, data mining, data sciences, knowledge representation and inference, semantic web and related topics. I am very interested in developing and teaching courses that incorporate significant research advances in the relevant disciplines and prepare students to address new research challenges (e.g. information integration, e-science) in computer science, data analytics, bioinformatics, and related areas.

I am also interested in offering focused special topics or seminar courses aimed at students in the undergraduate honors program, and graduate students on topics of current interest.

I would also enjoy contributing to the development of new graduate and undergraduate programs designed to train a new generation of scientists well-versed in computer and information sciences and the creative and skillful application of information processing approaches to address fundamental scientific problems in biological, agricultural, cognitive, environmental, health, physical and social sciences.

Given an opportunity, I would especially enjoy developing and teaching a course that introduces computer science as an intellectual discipline to a broad audience of undergraduates (and perhaps graduate students) drawn from across all areas of science, engineering, and the humanities.

Brief Description of Representative Courses Developed and Taught

- **Principles of Causal Inference.** Graduate course developed and taught at Pennsylvania State University. Causation versus association. Potential outcomes framework for counterfactual inference. Estimation of causal effects from observational data. Causal effects identifiability. Structural causal models. Confounding. Learning causal models from data. Causal effects transportability and related topics.
- **Principles of Artificial Intelligence**. Graduate Course taught at Iowa State University and Pennsylvania State University. Specification, design, implementation, and selected applications of intelligent software agents and multi-agent systems. Computational models of intelligent behavior, including problem solving, knowledge representation, reasoning, planning, decision making, learning, perception, action, communication and interaction. Reactive, deliberative, rational, adaptive, learning and communicative agents. Artificial intelligence programming. Graduate credit requires a research project and a written report. Oral and written reports.

- **Foundations of Informatics**. Graduate course developed and taught at Pennsylvania State University. The world viewed through an algorithmic lens. Review of theories of computation, complexity, information, and communication. Algorithmic abstractions of intelligence; algorithmic abstractions in cognitive and brains sciences; algorithmic abstractions in life sciences; algorithmic abstractions in the social sciences.
- **Principles of Machine Learning**. Graduate course taught at Iowa State University and Pennsylvania State University. Algorithmic models of learning. Design, analysis, implementation and applications of learning algorithms. Learning of concepts, classification rules, functions, relations, grammars, probability distributions, value functions, models, skills, behaviors and programs. Agents that learn from observation, examples, instruction, induction, deduction, reinforcement and interaction. Computational learning theory. Data mining and knowledge discovery using artificial neural networks, support vector machines, decision trees, Bayesian networks, association rules, dimensionality reduction, feature selection and visualization. Learning from heterogeneous, distributed, dynamic data and knowledge sources. Learning in multi-agent systems. Selected applications in automated knowledge acquisition, pattern recognition, program synthesis, bioinformatics and Internet-based information systems. Oral and written reports.
- **Computational Functional Genomics and Systems Biology**. Graduate course taught at Iowa State University. Algorithmic and statistical approaches in computational functional genomics and systems biology. Analysis of high throughput gene expression, proteomics, and other datasets obtained using system-wide measurements. Topological analysis, module discovery, and comparative analysis of gene and protein networks. Modeling, analysis, simulation and inference of transcriptional regulatory modules and networks, protein-protein interaction networks, metabolic networks, cells and systems: Dynamic systems, Boolean, and probabilistic models. Ontology-driven, network based, and probabilistic approaches to information integration.
- **Intelligent Multiagent Systems.** Graduate course taught at Iowa State University. Specification, design, implementation, and applications of multi-agent systems. Intelligent agent architectures; infrastructures, languages and tools for design and implementation of distributed multi-agent systems; Multi-agent organizations, communication, interaction, cooperation, team formation, negotiation, competition, and learning. Selected topics in decision theory, game theory, contract theory, bargaining theory, auction theory, and organizational theory. Selected topics in knowledge representation and ontologies. Agent-based systems and the Semantic Web. Applications in distributed intelligent information networks for information retrieval, information integration, inference, and discovery from heterogeneous, autonomous, distributed, dynamic information sources.
- **Advanced Topics in Computational Models of Learning.** Graduate. Selected topics in Computational Learning Theory (PAC learning, Sample complexity, VC Dimension, Occam Learning, Boosting, active learning, Kolomogorov Complexity, Learning under helpful distributions, Mistake Bound Analysis). Selected topics in Bayesian and Information Theoretic Models (ML, MAP, MDL, MML). Advanced statistical methods for machine learning. Selected topics in reinforcement learning.
- **Advanced Topics in Computational Intelligence.** Graduate. Advanced applications of artificial intelligence in bioinformatics, distributed intelligent information networks and the Semantic Web. Selected topics in distributed learning, incremental learning, multi-task learning, multi-strategy learning; Graphical models, multi-relational learning, and causal inference; statistical natural language processing; modeling the internet and the web; automated scientific discovery; neural and cognitive modeling.
- **Artificial Intelligence Graduate Research Seminar** Topics vary. Student-led discussion of research articles of current interest. Recent topics have included advanced topics in machine

learning and knowledge representation. Offered every semester.

- **Principles of Artificial Intelligence**. Undergraduate course taught at Iowa State University. Specification, design, implementation, and selected applications of intelligent software agents and multi-agent systems. Computational models of intelligent behavior, including problem solving, knowledge representation, reasoning, planning, decision making, learning, perception, action, communication and interaction. Reactive, deliberative, rational, adaptive, learning and communicative agents. Artificial intelligence programming. Graduate credit requires a research project and a written report. Oral and written reports.
- **Elements of Neural Computation**. Undergraduate course taught at Iowa State University. Introduction to theory and applications of neural computation and computational neuroscience. Computational models of neurons and networks of neurons. Neural architectures for associative memory, knowledge representation, inference, pattern classification, function approximation, stochastic search, decision making, and behavior. Neural architectures and algorithms for learning including perceptions, support vector machines, kernel methods, 86pprox.86 learning, instance based learning, reinforcement learning, unsupervised learning, and related techniques. Applications in Artificial Intelligence and cognitive and neural modeling. Hands-on experience is emphasized through the use of simulation tools and laboratory projects. Oral and written reports.

CONSULTING

Scientific consulting on data mining, artificial intelligence, bioinformatics, semantic technologies, knowledge-based systems, information integration, and information technology for several corporations, startups, and government organizations.

OTHER PROFESSIONAL ACTIVITIES

(Not including memberships in editorships of journals and service on standing review panels and study sections e.g., NIH)

ADVISORY BOARDS AND WORKING GROUPS

2019-	Eastern Regional Network
2020	MDEpinet Taskforce on AI and Blockchain
2016	Transatlantic Data Science Workshop
2015-	Electorate Nominating Committee, Information, Computing& Communication, AAAS
2016	IMA Workshop on Transdisciplinary Foundations of Data Science
2016-	mProv, Data Provenance for Mobile Health Data MD2K Center, Univ. of Memphis
2015-	Databrary, A Video Data Library for Developmental Science
2016-	Posed2, Inc
2015-	BioSNTR, an NSF EPSCOR Center at the University of South Dakota
2015-17	Chair, CCC Task Force on Convergence of Data and Computing
2016-17	CCC Task Force on Artificial Intelligence
2014-16	CCC Task Force on Health IT
2014	Masters Program in Data Science, Rochester Institute of Technology
2015	NSF INFEWS Workshop
2012	NSF Discovery Informatics Workshop
2012	NSF Knowledge Representation Workshop
2012	NSF Workshop on Population Health Measurement
2012	NSF Workshop on Next Generation Financial Cyberinfrastructure
2011	mHealth Evidence Workshop, NIH
2011-	Board of Directors, ACM Special Interest Group on Bioinformatics
2012-13	Interagency Working Group on Multi-Scale Modeling

2010	IciS Workshop on Integrating, Representing, and Reasoning with Human Knowledge, Snowbird, Utah
2010	AFOSFR Workshop on Dynamic Data-Driven Application Systems, Arlington, VA
2005-	External Advisory Committee, NSF HBCU-UP Program, Tuskegee University.
1999-03	Member of Advisory Board, Emergent Computational Neural Network Architectures, Universities of York, Edinburgh, and Sunderland, United Kingdom
2004	Member of Bioinformatics Group, National Science Foundation Intelligent Data Management PI Workshop, Boston, MA
2003	Member of Information Integration Working Group, National Science Foundation Intelligent Data Management PI Workshop, Seattle, WA.
1999	Member of Intelligent Agents Working Group, Information Institute, Information Directorate, Air Force Research Labs, Rome, New York
1999	Member of Search Committee, Senior Science and Technology Position in Defensive Information Warfare, Air Force Research Laboratory, Rome, New York.
2000	Proposal for a Ph.D. Program in Artificial Intelligence, University of Georgia Board of Regents
1999	Member, Intelligent Agents Working Group, Information Institute, Information Directorate, Air Force Research Labs
1998	Member, Intelligent Knowledge-Based Systems Working Group, Information Institute, Information Directorate, Air Force Research Labs
1997	Invited Participant, NSF Workshop on Decision Based Design, Sacramento, CA

CONFERENCE OR WORKSHOP PROGRAM CHAIR

2022	Program Chair	AAAI Conference on Artificial Intelligence
2021	Area Chair	AAAI Conference on Artificial Intelligence
2020	Area Chair	AAAI Conference on Artificial Intelligence
2019	Area Chair	AAAI Conference on Artificial Intelligence
2018	Co-Chair	ACM SIGSOFT International Workshop on Automated Specification Inference
2016	Chair	AAAI/CCC Workshop on Accelerating Science: A Grand Challenge for AI
2015	Chair	PSB Workshop on Discovery Informatics
2014	Co-chair	IEEE International Conference on Big Data
2014	Program Co-Chair	IEEE Conference on Bio and Medical Informatics
2014	Tutorial Program Chair	ACM Conference on Bioinformatics and Computational Biology
2014	Co-chair	ACM SIGKDD Workshop on Discovery Informatics
2014	Funding Co-chair	AAAI
2013	Tutorial Program Chair	ACM Conference on Bioinformatics and Computational Biology
2012	Workshop Program Chair	ACM Conference on Bioinformatics and Computational Biology
2011	Co-Chair	ACM Immunoinformatics and Computational Immunology Workshop
2011	Workshop Program Chair	ACM Conference on Bioinformatics and Computational Biology

2011	Co-Chair	ACM Immunoinformatics & Comp. Immunology Workshop
2010	Co-Chair	ACM Immunoinformatics & Comp. Immunology Workshop
2010	Area Chair	IEEE Conference on Tools with Artificial Intelligence
2009	Sponsorship co-chair	SIAM Conference on Data Mining
2009	Area Chair	IEEE Conference on Tools with Artificial Intelligence
2009	Organizer and Chair	Doctoral Student Forum, SIAM Conference on Data Mining
2008	Vice Program Chair	International Semantic Web Conference
2008	Track Chair	SIAM Conference on Data Mining
2007	Organizer and Co-Chair	IJCAI Workshop on Semantic Web for Collaborative Knowledge Acquisition, IJCAI-2007 Hyderabad, India
2006	Organizer and Co-Chair	First International Workshop on Modular Ontologies, International Semantic Web Conference, Athens, GA
2006	Organizer and Co-Chair	AAAI Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition (SWeCKa 2006), Washington, DC.
2005	Organizer and Chair	IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Information Sources, IEEE Conference on Data Mining, Houston, Texas
2004	Program Co-Chair	International Conference on Intelligent Information Processing, Beijing, China
2003	Organizer and Chair	Computational Intelligence Workshop John Vincent Atanasoff Symposium on Advanced Computing, Iowa State University, Ames, Iowa
2002	Program Chair	Conference on Computational Biology and Genome Informatics, Durham, North Carolina
2002	Track Chair	Artificial Life, Agents, and Adaptive Behavior Genetic and Evolutionary Computing Conference, New York
2001	Organizer and Chair	Workshop on Knowledge Discovery from Heterogeneous, Distributed, Dynamic, Autonomous Data and Knowledge Sources. International Joint Conference on Artificial Intelligence, Seattle.
2001	Organizer and Co-Chair	Workshop on Cognitive Agents and Inter-agent Interaction, International Conference on Cognitive Science, Beijing
2000	Organizer and Co-Chair	Workshop on Learning from Sequential and Temporal Data, International Conference on Machine Learning, Palo Alto, CA
1999	Organizer and Co-Chair	Workshop on Computation with Neural Systems National Conference on Artificial Intelligence (AAAI), Orlando
1999	Track Chair	Artificial Life, Agents, and Adaptive Behavior Genetic and Evolutionary Computing Conference, Orlando
1998	Program Chair	International Colloquium on Grammatical Inference Ames, Iowa
1997	Organizer and Co-Chair	Workshop on Automata Induction, Grammatical Inference, and Language Acquisition International Conference on Machine Learning, Nashville
1992	Organizer and Chair	Workshop on Symbolic and Subsymbolic Information Processing, Neural Circuits and Systems, Conference on Neural Information Processing, Vail, CO

Conference Advisory Committee Membership

2008	Workshop on Web Service Composition and Adaptation, WSCA 2008
2007	Computational Structural Bioinformatics Workshop, BIBE 2007
2004	International Conference on Intelligent Knowledge Systems, Turkey
1997	International Conference on Evolutionary Computation
1997	International Conference on Artificial Intelligence, Cairo, Egypt

Conference or Workshop Program Committee Membership

2020	ACM/IMS Conference on the Foundations of Data Science
2019	PC Member, International Joint Conference on Artificial Intelligence
2018	PC Member, International Conference on Machine Learning
2017	SPC Member, AAAI Conference on Artificial Intelligence
2017	ACM Conference on Bioinformatics and Computational Biology
2016	AAAI Conference on Artificial Intelligence
2016	IEEE Conference on Bioinformatics and Biomedicine
2016	Conference on Data Integration in the Life Sciences
2016	ACM Conference on Bioinformatics and Computational Biology
2016	Data Science, Learning, and Applications to Biomedical & Health Sciences
2016	SPC Member, AAAI Conference on Artificial Intelligence
2015	PC Member, AAAI Conference on Artificial Intelligence
2015	PC Member, IEEE Conference on Bioinformatics and Biomedicine
2015	PC Member, Conference on Data Integration in the Life Sciences
2015	PC Member, ACM Conference on Bioinformatics and Computational Biology
2015	PC Member, AAAI Conference on Artificial Intelligence
2014	PC Member, AAAI Conference on Artificial Intelligence
2014	IEEE Big Data Congress
2014	ACM Conference on Bioinformatics, Computational Biology, and Health Informatics
2013	Conference on Data Integration in Life Sciences (DILS 2013)
2013	IEEE International Conference on Big Data
2013	IEEE Big Data Congress
2013	IEEE Conference on Bioinformatics and Biomedicine (BIBM 2013)
2012	International Semantic Web Conference (ISWC 2012)
2012	AAAI Conference on Artificial Intelligence (AAAI 2012)
2012	Conference on Data Integration in Life Sciences (DILS 2012)
2012	IEEE Conference on Tools with Artificial Intelligence (ICTAI 2012)
2012	International Conference on Bioinformatics Models, Methods, and Algorithms
2012	International Conference on Health Informatics (IHI 2012)
2012	IEEE Conference on Biomedical Computing 2012
2012	SIAM Data Mining Conference (SDM 2012)
2011	SIAM Data Mining Conference (SDM 2011)
2011	ACM Conference on Bioinformatics and Computational Biology (ACM-BCB 2011)
2011	Intelligent Systems in Computational Biology (ISMB / ECCB 2011)
2011	Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2011)
2011	IEEE Conference on Tools with Artificial Intelligence (ICTAI 2011)
2011	ACM/WIC/IEEE Conference on Intelligent Agent Technology (IAT 2011)
2011	ACM International Health Informatics Symposium (IHI 2011)
2010	National Conference on Artificial Intelligence (AAAI 2010)

2010	International Conference on Machine Learning (ICML 2010)
2010	IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2010)
2010	ACM International Health Informatics Symposium (IHI 2010)
2010	ACM/WIC/IEEE Conference on Intelligent Agent Technology (IAT 2010)
2010	ACM Conference on Bioinformatics and Computational Biology (ACM-BCB 2010)
2010	International Symposium on Bioinformatics Research and Applications (ISBRA 2010)
2010	Workshop on Modular Ontologies (WOMO 2010)
2010	SIAM Data Mining Conference (SDM 2010)
2010	Pacific Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2010)
2010	International Conference on Advances in Semantic Processing (SEMAPRO 2010)
2009	Intelligent Systems in Molecular Biology (ISMB 2009)
2009	ACM/WIC/IEEE Conference on Intelligent Agent Technology (IAT 2009)
2009	IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2009)
2009	International Conference on Intelligent Data Engineering and Automated Learning
2009	International Symposium on Bioinformatics Research and Applications (ISBRA 2009)
2009	International Conference on Advances in Semantic Processing (SEMAPRO 2009)
2008	International Conference on Machine Learning (ICML 2008)
2008	ACM/WIC/IEEE Conference on Intelligent Agent Technology (IAT 2008)
2008	SIAM Conference on Data Mining (SDM 2008)
2008	International Colloquium on Grammatical Inference (ICGI 2008)
2008	International Symposium on Bioinformatics Research and Applications (ISBRA 2008)
2007	ACM SIGKDD Conference on Data Mining and Knowledge Discovery (KDD 2007)
2007	Intelligent Systems in Molecular Biology (ISMB 2007)
2007	International Colloquium on Grammatical Inference (ICGI 2007)
2007	ACM/WIC/IEEE Conference on Intelligent Agent Technology (IAT 2007)
2007	IEEE Symposium on Computational Intelligence and Data Mining (CIDM 2007)
2007	AAAI Workshop on Semantic e-Science (SeS 2007)
2007	2nd International Workshop on Modular Ontologies (WoMO 2007)
2007	ICWS Workshop on Service Composition and Adaptation (WSCA 2007)
2007	IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2007)
2007	IEEE International Conference on Bioinformatics and Bioengineering (BIBE 2007)
2006	International Conference on Machine Learning (ICML 2006)
2006	AAAI Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition (SWeCKa 2006)
2006	First International Workshop on Modular Ontologies (WoMO 2006)
2006	IEEE Conference on Tools with Artificial Intelligence (ICTAI 2006)
2006	International Colloquium on Grammatical Inference (ICGI 2006)
2006	ACM / IEEE / WIC Conference on Web Intelligence (WI 2006)
2006	International Workshop on Algorithms in Bioinformatics (WABI 2006)
2006	8th International Conference on Data Warehousing and Knowledge Discovery
2006	IASTED International Conference on Computational Intelligence (CI 2006)
2006	IEEE Conference on Granular Computing (IEEE-GrC 2006)
2005	IEEE International Conference on Data Mining (ICDM 2005)
2005	International Conference on Algorithmic Learning Theory (ALT 2005)
2005	IEEE Conference on Tools with Artificial Intelligence (ICTAI 2005)
2005	International Conference on Machine Learning (ICML 2005)
2005	IEEE/ACM Conference on Intelligent Agent Technology (IAT 2005)
2005	ACM SIGKDD Workshop on Data Mining in Bioinformatics (BIOKDD 2005)

2005	Indian International Conference on Artificial Intelligence (IICAI 2005)
2004	IEEE International Conference on Data Mining (ICDM 2004)
2004	International Conference on Machine Learning (ICML 2004)
2004	IEEE/WIC/ACM Conference on Intelligent Agent Technology (IAT 2004)
2004	AAAI Workshop on Semantic Web Personalization
2004	IEEE Conference on Bioinformatics and Bioengineering (BIBE 2004)
2004	International Colloquium on Grammatical Inference (ICGI 2004)
2004	SIAM Bioinformatics Workshop, SIAM International Conference on Data Mining (SDM 04)
2004	IEEE Conference on Tools with Artificial Intelligence (ICTAI 2004)
2004	Midwestern Conference on Artificial Intelligence and Cognitive Science (MAICS 2004)
2003	IEEE International Conference on Data Mining (ICDM 2003)
2003	IEEE Conference on Tools with Artificial Intelligence (ICTAI 2003)
2003	International Conference on Intelligent Systems Design and Applications (ISDA 2003)
2003	Midwestern Conference on Artificial Intelligence and Cognitive Science (MAICS 2003)
2002	International Colloquium on Grammatical Inference (ICGI 2002)
2002	International Conference on Hybrid Intelligent Systems (HIS 2002)
2002	International Conference on Intelligent Systems Design and Applications (ISDA 2002)
2002	Midwestern Conference on Artificial Intelligence and Cognitive Science (MAICS 2002)
2002	Network Applications in Bioinformatics (NETTAB) Workshop at Bologna, Italy, 2002.
2002	Fourth International Workshop in Frontiers of Evolutionary Computation (FEA-2002)
2001	International Conference on Machine Learning (ICML 2001)
2001	International Symposium on Artificial Intelligence, India (ISAI 2001)
2001	International Joint Conference on Neural Networks (IJCNN 2001)
2001	International Conference on Hybrid Intelligent Systems (HIS 2001)
2001	SIAM Workshop on Mining Scientific Data Sets (at SDM 2001)
2001	Midwestern Conference on Artificial Intelligence and Cognitive Science (MAICS 2001)
2001	Genetic and Evolutionary Computing Conference (GECCO 2001)
2001	Workshop on Gene Expression. Genetic and Evolutionary Computing Conference, 2001
2000	International Colloquium on Grammatical Inference (ICGI 2000)
2000	Genetic and Evolutionary Computing Conference (GECCO 2000)
2000	Midwestern Conference on Artificial Intelligence and Cognitive Science (MAICS 2000)
1999	National Conference on Artificial Intelligence (AAAI 1999)
1999	Midwestern Conference on Artificial Intelligence and Cognitive Science (MAICS 1999)
1998	International Colloquium on Grammatical Inference (ICGI 1998)
1998	Genetic Programming Conference (GP 1998)
1997	International Conference on Machine Learning (ICML 1997)
1997	Genetic Programming Conference (GP 1997)
1996	Genetic Programming Conference (GP 1996)
1997	Midwest Artificial Intelligence and Cognitive Science Conference (MAICS 1997)
1996	World Congress on Neural Networks (WCNN 1996)
1996	Midwest Artificial Intelligence and Cognitive Science Conference (MAICS 1996)
1995	World Congress on Neural Networks (WCNN 1995)
1995	Midwest Artificial Intelligence and Cognitive Science Conference (MAICS 1995)
1993	International Simulation Technology Conference (SIMTEC 1993)
1993	University of New Brunswick Artificial Intelligence Symposium
1992	International Simulation Technology Conference (SIMTEC 1992)

Journal Referee

Applied Intelligence
Bioinformatics
BMC Bioinformatics
BMC Cancer
Connection Science
Genetic Programming
IEEE Computer
IEEE Expert
IEEE Intelligent Systems
IEEE Transactions on Data and Knowledge Engineering
IEEE Transactions on Evolutionary Computation
IEEE Transactions on Neural Networks
IEEE Transactions on Pattern Analysis and Machine Intelligence
IEEE Transactions on System, Man, and Cybernetics
International Journal of Data Mining and Bioinformatics
Information and Computation
Information Fusion
Information Sciences
Journal of Computational Biology
Journal of Machine Learning Research
Nucleic Acids Research
Neural Computation
Neural Networks
Machine Learning
Pattern Recognition
Proteins Structure, Function and Bioinformatics

Proposal Reviewer

National Science Foundation, USA
National Institutes of Health, USA
US Civilian Research Development Foundation, USA
United States Department of Agriculture, USA
Dutch National Science Foundation, Netherlands
European Physical Science and Engineering Research Council, United Kingdom
Irish National Science Foundation, Ireland
National Science and Engineering Research Council, Canada

Other Professional Service

Book and/or journal proposals for MIT Press, Academic Press, CRC Press, and Springer Verlag.
External Referee, Promotion and Tenure.

DEPARTMENTAL, COLLEGE, AND UNIVERSITY SERVICE

2020-	Member AI Consortium Coordination Committee, ICDS	PSU
2020-	Member, Data Sciences Coordination Committee	PSU
2015-	Director, Center for Big Data Analytics and Discovery Informatics	PSU
2014-	Associate Director, Institute for Computational and Data Sciences	PSU
2017-	Member, University Health Sciences Council	PSU

2019	AI in Medicine Symposium Organizer and Chair		PSU
2019-	Member, EHR research data governance committee		PSU
2019-	Member, Faculty Council, Institute for Computational and Data Sciences		PSU
2019-20	Member, Data Sciences Faculty Search Committee		PSU
2019-20	Member, Education and Data Science Faculty Search Committee		
2018-19	Chair, Biomedical Data Sciences Faculty Search Committee		PSU
2018-19	Member, Astroinformatics Faculty Search Committee		PSU
2018-19	Member, Geoinformatics and Remote Sensing Faculty Search Committee		PSU
2017-	Member, Faculty Council, College of IST		PSU
2018-19	Member, Informatics Graduate Program Advisory Committee		PSU
2018-	Member, Cognitive and Immersive Technologies Working Group		PSU
2016-17	Chair, Data Science faculty Search Committee, College of IST		PSU
2016-17	Faculty Council, College of IST		PSU
2016-17	Graduate Recruitment Committee		PSU
2016-	Executive Committee, Biomedical Data Sciences Training Program		PSU
2014-	Penn State Clinical and Translational Sciences Institute		PSU
2016-	Microbiome Center Executive Committee		PSU
2016-17	Promotion and Tenure Committee, College of IST		PSU
2016-	Data Sciences Program Supervisory Committee		PSU
2015-17	Research Advisory Committee, College of IST		PSU
2015-	Institute for Cyberscience Coordinating Committee		PSU
2015-16	Data Sciences Undergraduate Program Task Force		PSU
2016-	Research Data and Computing Committee, VPR		PSU
2015-16	Data Sciences Faculty Search Committee, College of IST		PSU
2014-16	Graduate Advisory Committee, College of IST		PSU
2016-17	Lecturer Search Committee, College of IST		PSU
2015-17	Biomedical Engineering Search Committee		PSU
2013-14	Member, Vice Provost's Taskforce on Computing and Data Infrastructure		PSU
2013-14	Member, Strategic Planning Committee, College of Information Science and Technology		PSU
2005-2013	Director, Center for Computational Intelligence, Learning and Discovery		ISU
2010-2011	Member, Supervisory Committee, Bioinformatics and Computational Biology Graduate Program		ISU
2009-2010	Director of Research	Computer Science	ISU
2009-2010	Director, Center for Computational Intelligence, Learning and Discovery		ISU
2009-2010	Member, Research Computing Council		ISU
2009-2010	Chair, Promotion & Tenure Steering Committee	Computer Science	ISU
2009-2010	Member, Supervisory Committee, Bioinformatics and Computational Biology Undergraduate Program		ISU
2009-2010	Member, Supervisory Committee, Bioinformatics and Computational Biology Graduate Program		ISU
2009-2010	Director of Research	Computer Science	ISU
2008-2009	Member, Research Computing Council		ISU
2008-2009	Member, e-Science Steering Committee		ISU
2008-2009	Member, Promotion & Tenure Steering Committee	Computer Science	ISU

2008-2009	Member, Supervisory Committee, Bioinformatics and Computational Biology Undergraduate Program		ISU
2008-2009	Member, Supervisory Committee, Bioinformatics and Computational Biology Graduate Program		ISU
2007-2008	Member, Supervisory Committee, Bioinformatics and Computational Biology Undergraduate Program		ISU
2007-2008	Member, Research Computing Council		ISU
2007-2008	Member, Systems Biology Steering Committee		ISU
2007-2008	Member, Graduate Committee	Computer Science	ISU
2007-2008	Member, Promotion & Tenure Steering Committee	Computer Science	ISU
2007-2008	Member, Graduate Admissions Committee	Computer Science	ISU
2007-2008	Member, Faculty Search Committee	Computer Science	ISU
2007-2008	Member, Professional and Scientific Staff Hiring Committee (ad hoc)	Computer Science	ISU
2007-2008	Director of Research	Computer Science	ISU
2006-2007	Member, Graduate Committee	Computer Science	ISU
2006-2007	Chair, Research Infrastructure Committee	Computer Science	ISU
2006-2007	Director of Research	Computer Science	ISU
2006-2007	Director, Center for Computational Intelligence, Learning and Discovery	ISU	ISU
2005-2006	Chair, Graduate Admissions Committee	Computer Science	ISU
2005-2006	Member, Departmental Web Committee	Computer Science	ISU
2005-2007	Ex Officio Member, Supervisory Committee, Bioinformatics & Computational Biology Graduate Program		ISU
2005-2007	Member, Planning Committee, Bioinformatics Undergraduate Program		ISU
2005-2006	Director of Research	Computer Science	ISU
2005-2006	Director, Center for Computational Intelligence, Learning and Discovery	ISU	ISU
2005-2006	Member, Promotion and Tenure Steering Committee	Computer Science	ISU
2004-2005	Member, Graduate Committee	Computer Science	ISU
2004-2005	Committee for review of the Office of the Vice President of Business & Finance		ISU
2003-2004	Member, ISU Information Technology Working Group: Research		ISU
2003-2004	Member, Graduate Admissions Committee	Computer Science	ISU
2002-2003	Member, Graduate Admissions Committee	Computer Science	ISU
2003-2004	Member, Provost's Information Technology Advisory Group		ISU
2002-2003	Member, Human Computer Interaction Graduate Program Steering Committee		ISU
2000-2001	Member, Department Chair Search Committee	Computer Science	ISU
1999-	Member, Supervisory Committee, Interdepartmental Bioinformatics & Computational Biology Graduate Program		ISU
1999	Coordinator, Complex Adaptive Systems Workshop		ISU
1999-2002	Member, Supervisory Committee, Complex Adaptive Systems Graduate Minor		ISU
1998-1999	Member, Graduate Committee	Computer Science	ISU

1998	Departmental Delegate, Annual Faculty Conference	Computer Science	ISU
1998	Member, Advisory Committee to Department Chair, Presidential Taskforce on Information Technology	Computer Science	ISU
1997-1998	Member, Graduate Committee	Computer Science	ISU
1996-1997	Member, Carver Trust Grant Proposal Review Committee		ISU
1997-1999	Member, Bioinformatics & Computational Biology Program Steering Committee		ISU
1996-	Member, Promotion and Tenure Committee	Computer Science	ISU
1996-1997	Member, Graduate Committee	Computer Science	ISU
1996-1997	Coordinator, Graduate Student Orientation	Computer Science	ISU
1996	Member, Industry Day Organization Committee	Computer Science	ISU
1996	Advisor, Iowa State University Student Team, AAAI Robot Competition		ISU
1996-1997	Member, Carver Trust Grant Proposal Review Committee		ISU
1995-1996	Faculty Secretary	Computer Science	ISU
1995-1999	Member, Interdepartmental Neuroscience Graduate Program Supervisory Committee		ISU
1994-1996	Member, Departmental Strategic Planning Committee	Computer Science	ISU
1990-1996	Member, Graduate Admissions Committee	Computer Science	ISU
1992-1994	Member, Liberal Arts and Sciences Honors Program Committee		ISU

References available upon request