BIOGRAPHICAL SKETCH

Provide the following information for the Senior/key personnel and other significant contributors.
Follow this format for each person. **DO NOT EXCEED FIVE PAGES.**

NAME: Xia Jiang

eRA COMMONS USER NAME (credential, e.g., agency login): XIAJIANG

POSITION TITLE: Associate Professor, Department of Biomedical Informatics, University of Pittsburgh, School of Medicine

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, include postdoctoral training and residency training if applicable. Add/delete rows as necessary.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	Completion Date MM/YYYY	FIELD OF STUDY
Southern Institute of Metallurgy, P. R. China	BS	1989	Metallurgy
Rose-Hulman Institute of Technology, Terre Haute, IN	MS	1997	Mechanical Engineering
Northeastern Illinois University, Chicago, IL	MS	1999	Computer Science
University of Pittsburgh, Pittsburgh, PA	PHD	2008	Biomedical Informatics

A. Personal Statement

I am an Associate Professor in the Department of Biomedical Informatics and the Intelligent Systems Program at the University of Pittsburgh, and in the joint CMU-Pitt Computational and Systems Biology PhD program. My overall research direction is to take computational and informatics approach to help solve biomedical problems. My areas of research interest include Bayesian networks modeling and learning, machine learning, biomarker detection, biomedical prediction, decision analysis and modeling, and cancer informatics such as cancer signal transduction pathway (STP) analysis. I am the pioneering researcher in applying Bayesian network scoring to learning epistatic interactions from GWAS data. In that regard, I defined a class of Bayesian network models for learning epistasis called SNP patterns. I developed a novel score criterion for detecting epistatic biomarkers, named as the Bayesian Network Minimum Bit Length (BNMBL) score, and I identified a bound for the Bayesian score of a SNP pattern. I also developed several novel and efficient epistasis learning and evaluation algorithms including Multiple Beam Search (MBS), Repeated Extended Greedy Approach to Learning (REGAL), Bayesian Network Posterior Probability (BNPP), the Interaction Strength (IS), and Information Gain (IGain) based algorithms for learning interactions. These algorithms have not only been applied to the detection of epistatic biomarkers, including rare variants associated with a disease, but also have been used to detect the interaction between microRNA and mRNA in breast cancer. Most recently, I developed a novel algorithm called Direct Causal Learner (DCL) which can be used to learn direct causes of a clinical outcome. Our experiments demonstrated that DCL outperformed some of the most well-known existing causal learning algorithms.
My research has resulted in 58 peer-reviewed publications, 41 of which are related to the development and evaluation of new algorithms and informatics tools. I am the contact PI of the DOD BCRP Level II (W81XWH1910495) grant about using deep Learning to develop prediction models for breast cancer metastasis. I was the contact PI of an NLM K99/R00 project (LM010822) that concentrated on efficient epistasis learning, and the contact PI of an NLM R01 project (LM011663) about breast cancer clinical decision support. These two projects have resulted in four inventions copy-righted by the University of Pittsburgh Innovation Institute. Via these projects, I successfully established a collaboration with breast cancer physicians including Dr. Adam Brufsky (a co-I of both the current and the proposed DoD project) at UPCI and Dr. Seema Khan at the Northwestern University. I have also established a collaboration with Dr. Alan Wells (a co-I of both the current and the proposed DoD project) during the past eight years due to a shared interest in breast cancer

metastasis. Finally, I am a participating co-I on two projects concerning big data science and causal modeling and learning (the BD2K CCMD project and the state-sponsored BD4BH project).

B. Positions and Honors

Employment

7/1989-12/1991	Responsible Editor of China National Tungsten Industry, Division of Tungsten and Molybdenum Research, China National Non-ferrous Metals Industry Corporation, Beijing, China
12/1991-08/1995	Researcher, Division of Technology and Economy Research, Institute of Technology and Economy Research, China National Non-ferrous Metals Industry Corporation, Beijing, China
08/1995-04/1997	Part-time Teaching Assistant, Rose-Hulman Institute of Technology
04/1997-04/2002	Software Engineer, Warehouse Equipment Inc., Elk Grove Village, IL
05/2002-08/2005	Instructor, Department of Computer Science, Northeastern Illinois University
08/2005-11/2008	Graduate Student Researcher, Department of Biomedical Informatics, University of Pittsburgh
12/2008-9//2010	Postdoctoral Scholar, Department of Biomedical Informatics, University of Pittsburgh
10/2010-12/2011	Postdoctoral Associate, Department of Biomedical Informatics, University of Pittsburgh
01/2012-10/2016	Assistant Professor, Department of Biomedical Informatics, University of Pittsburgh
11/2017-present	Associate Professor, Department of Biomedical Informatics, University of Pittsburgh

Honors

1986	"Excellent San-hao Student" scholarship, Southern Institute of Metallurgy
1987	1st place in "Poem Recitation Contest", Southern Institute of Metallurgy
1988	2nd place in "English Contest", Southern Institute of Metallurgy
1988	Winner of "Academic Excellence" award, Southern Institute of Metallurgy
1993	2nd place, "Annual Excellent Research" award for research project *Investigation and Prediction of Lead-Zinc Market*, Institute of Technology and Economy Research, China Non-ferrous Metals Industry Corporation, Beijing, China
2008	Semi-finalist of "the Best Publication of the Year" award for the *Journal of Biomedical Informatics* paper "Bayesian Prediction of an Epidemic Curve", International Society for Disease Surveillance Annual Conference
2008	National Library of Medicine Postdoctoral Training Scholarship
2009	National Library of Medicine Postdoctoral Training Scholarship
2010	National Library of Medicine Postdoctoral Training Scholarship
2010	Finalist Award, student paper competition at American Medical Informatics Association (AMIA) 2010 Annual Symposium for paper titled "A Fast Algorithm for Learning Epistatic Genomic Relationships", Proceedings of American Medical Informatics Association (AMIA) 2010 Annual Symposium

Professional Societies

2006-Present	Member, Association for the Advancement of Artificial Intelligence (AAAI)
2007-2008, 2012	Member, Association for Computing Machinery (ACM)
2007-Present	Member, American Medical Informatics Association (AMIA)

C. Contribution to Science

1. Learning epistasis, interaction, and causation from data: The interaction of entities/attributes is a universal phenomenon and plays a critical role in many aspects of human life. For example, it is believed that interactions among genetic loci (epistasis) may contribute significantly to susceptibility to common diseases. However, learning interactions from data is an area of machine learning that has not yet been fully explored. I was the first

to apply Bayesian network scoring to learning epistatic interactions from GWAS data. Specifically, I defined a class of Bayesian Network models for learning epistasis called SNP patterns, and I developed the Bayesian Network Minimum Bit Length (BNMBL) score, a scoring criterion for detecting epistatic interactions that is based on the minimum description length principle [1]. I identified a bound for the score of a SNP pattern. The bound provides an upper limit on the Bayesian score of any pattern that could be obtained by expanding a given pattern. I developed MBS-IGain [2], which uses both Bayesian network scoring and information theory to learn interactions. MBS-IGain substantially out-performed 9 other methods, and may become the premier method for learning interactions. I developed evaluation measures for learning interactions, and a tool for ameliorating multiple hypotheses testing issues when learning from high-dimensional data. We developed a novel Bayesian network based causation learning algorithm that performed better than well-known existing causal learning algorithms in our evaluation [3]. I developed interaction learning methods by integrating information gain with Bayesion network learning and applied them to learning interactive causes of breaskt cancer and risk factors for breast cancer metastasis [4].

1) **Jiang X**, Barmada MM, Visweswaran S. Identifying genetic interactions in genome-wide data using Bayesian networks. Genetic Epidemiology 2010; 34(6): 575-81. PMID: 20568290. PMCID: PMC3931553.
2) **Jiang X**, Jao J, Neapolitan RE. Learning predictive interactions using information gain and Bayesian network scoring. PLoS ONE 2015; 10(12). PMCID: PMC4666609.
3) Rathnam, CS, Lee, S, and **Jiang X**, An algorithm for direct causal learning of influences on patient outcomes. Artificial Intelligence in Medicine, Jan 2017; DOI:10.1016/j.artmed.2016.10.003.
4) **Jiang, X**., Wells, A., Brufsky, A. et al. Leveraging Bayesian networks and information theory to learn risk factors for breast cancer metastasis. BMC Bioinformatics 21 (1–17), 298 (2020). https://doi.org/10.1186/s12859-020-03638-8

2. Developing Bayesian network based prediction and decision support models: Medical diagnosis, prognosis determination, and treatment selection require prediction and decision making. Prediction is a critical step in the effort to recommend decisions that maximize the expected utility of the outcomes to the patient. Clinical data are becoming increasingly available in electronic form, providing tremendous opportunities for developing accurate classification and prediction methods. My collaborators and I contributed to patient outcome prediction and decision modeling in the following two ways: 1) We developed new methods for predicting the survival of breast cancer patients using integrated heterogeneous data and Bayesian network modeling [1]; 2) We designed decision analytic models [2,3]; My other contributions to prediction are my earlier research studies pertaining to Bayesian network based disease outbreak modeling and detection. For example I developed methods for predicting disease outbreaks and epicurves [4].

1) **Jiang X**, Xue D, Brufsky AM, Khan SA, Neapolitan RE. A new method for predicting patient survivorship using efficient Bayesian network learning. Cancer Informatics 2014; 13 (2):47-57. PMID: 24558297. PMCID: PMC3928477.
2) Neapolitan R, **Jiang X**, Ladner DP, Kaplan B. A primer on Bayesian decision analysis with an application to a kidney transplant decision. Transplantation 2016; 100(3): 489-496. PMCID: PMC4818954.
3) **Jiang X**, Wells A, Brufsky A, Neapolitan RE, A Clinical Decision Support System Learned From Data to Personalize Treatment Recommendations Towards Preventing Breast Cancer Metastasis, PLoS One 14(3): e0213292, March 8, 2019; doi:10.1371/journal.pone.0213292
4) **Jiang X**, Wallstrom GL. A Bayesian network for outbreak detection and prediction. 21st Association for the Advancement of Artificial Intelligence (AAAI) Conference. 2006: 1155-60.

3. Machine learning and deep learning: We developed an efficient neural network based machine learning method for conducting biomedical prediction [1]. We applied machine learning methods to ubiquitin site prediction using physicochemical properties of protein segments, an effort that is related to the study of DNA repair mechanisms [2]. We developed deep feedforward neural network models for conducting personalized prediction of breast cancer metastasis, and we applied grid search to both deep learning and machine learning for predicting later occurence of breast cancer metastasis using clinical data [3,4].

1) Cai B, **Jiang X**. Novel Artificial Neural Network Method for Biomedical Prediction based on Matrix Pseudo-Inversion. Journal of Biomedical Informatics 2014; 48: 114-21. PMCID: PMC4004678.

2) Cai B, **Jiang X.** Computational methods for ubiquitination site prediction using physicochemical properties of protein sequences. BMC Bioinformatics 2016; 17:116. PubMed PMID: 26940649. PMCID: PMC4778322.
3) Gomez Marti J., Brufsky A, Wells A, **Jiang X**. Machine Learning of Discern Interactive Clusters of Risk Factor for Late Recurrence of Metastatic Breast Cancer. *Cancers (Basel)*. 2022 Jan 5;14(1):253. PMID: 35008417 PMCID: PMC8750735
4) **Jiang X**; Xu C. Deep Learning and Machine Learning with Grid Search to Predict Later Occurrence of Breast Cancer Metastasis Using Clinical Data. Journal of Clinical Medicine. 2022, 11, 5772. https://doi.org/10.3390/jcm11195772

4. Developing computational efficient search algorithms: A difficulty when mining interactions from high-dimensional datasets concerns the *curse of dimensionality*. With the arrival of the big data era, developing efficient search algorithms has become one of the most important and challenging tasks in the machine learning community. Spatial cluster detection consists of finding spatial subregions of some larger region where clusters of some event are occurring. I developed a recursive learning algorithm called *refine* for fast spatial cluster detection of complex sub-regions [1]. I designed a dynamic programming algorithm for spatial cluster detection by collaborating with my colleagues [2] . I developed efficient Bayesian network-based epistasis learning algorithms, namely Multiple Beam Search (MBS) and Repeated Extended Greedy Approach to Learning (REGAL) [3]. I extended this research by developing a learning tool, which uses the BDeu score and MBS to learn SNP interactions, and then uses the BNPP (an evaluation tool I developed) to evaluate the probability of that association. The complete system is called Learning and Evaluation Association Patterns (LEAP) [4]. I showed that LEAP is an effective tool for extracting candidate interacting causal patterns from high-dimensional datasets and determining their probability. Striving to improve further on the interaction learning power of LEAP.

1) **Jiang X**, Cooper GF. A recursive algorithm for spatial cluster detection. AMIA Annu Symp Proc 2007; 369-73. PMCID: PMC2655859.
2) Sverchkov Y, **Jiang X**, Cooper GF. Spatial cluster detection using dynamic programming. BMC Medical Informatics and Decision Making 2012; 12-22. PMCID: PMC3403878.
3) **Jiang X**, Neapolitan RE, Barmada MM, Visweswaran S, Cooper GF. A fast algorithm for learning epistatic genomic relationships. AMIA Annual Symposium; 2010:341-5
4) **Jiang X**, Neapolitan RE. LEAP: biomarker inference through learning and evaluating association patterns. Genetic Epidemiology 2015; 39(3):173–184. PMID: 25677188. PMCID: PMC4666609.

5. System biology:
In [1] with colleagues, I applied SPIA to 10 TCGA cancer datasets, and 157 KEGG STPs. We also performed a pan cancer analysis in which all 10 datasets were merged. We obtained results confirmed by the literature and new results. Specifically, in the pan cancer analysis the 4 most notable pathways learned are all known to be major players in cancer. I developed and applied new methodology for investigating the alteration of signal transduction pathways (STPs) in diseases. In [2] I developed a Bayesian network (BN) based method, called CASA, for learning aberrant STPs in disease from data, and applied it to infer causal melocular networks as part of a community effort. In [3,4] with colleagues I applied both CASA and SPIA (a non-BN based method for learning STP alterations from data) to a TCGA ovarian cancer dataset and 26 STPs, 20 of which are believed to be implicated in cancer.

1) Neapolitan R, Horvath CM, **Jiang X**. Pan-cancer analysis of TCGA data reveals notable signaling pathways. BMC Cancer; 2015; 15: 516.
2) Hill SM, Heiser LM, Cokelaer T,,, **Jiang X**,,,. Empirical Inferring causal molecular networks: empirical assessment through a community-based effort. Nature Methods 2016; 13: 310–318. PMID: 26901648. PMCID: PMC4854847.
3) Neapolitan R, **Jiang X**. Inferring aberrant signal transduction pathways in ovarian cancer from TCGA data. Cancer Informatics; 2014; Suppl. 1: 29-36.
4) Cai C, Cooper GF, Lu K, Ma X, …, **Jiang X**, … , Lu X, Systematic Discovery of the Functional Impact of Somatic Genome Alterations in Individual Tumors through Tumor-specific Causal Inference, PLoS Comput Biol 15(7): e1007088. July 5, 2019; https://doi.org/10.1371/journal.pcbi.1007088

Complete List of Published Work:
https://scholar.google.com/citations?hl=en&user=6Yid6tMAAAAJ