

Ankit Agrawal

CURRENT POSITION

Research Associate Professor

*Department of Electrical Engineering and Computer Science
McCormick School of Engineering and Applied Science
Northwestern University*

Office: M384, TECH
Technological Institute
2145 Sheridan Road
Evanston, Illinois 60208-3118

Ph: 847-491-8163
Fax: 847-491-4455
Email: ankitag@eecs.northwestern.edu

Website: <http://users.eecs.northwestern.edu/~ankitag>
Google Scholar: <http://scholar.google.com/citations?user=5e7JslgAAAAJ>



RESEARCH INTERESTS

Data Mining, High Performance Computing, Materials Informatics, Healthcare Informatics, Social Media Analytics.

EDUCATION

- Ph.D in [Computer Science](#) (with Minor in [Bioinformatics and Computational Biology](#)), [Iowa State University](#), USA. August 2006 - August 2009 (defended May 2009).
 - GPA: 4.0/4.0
 - Dissertation: Sequence-Specific Sequence Comparison Using Pairwise Statistical Significance
 - Advisor: [Xiaoqiu Huang](#)
- B.Tech. in [Computer Science and Engineering](#), [Indian Institute of Technology, Roorkee](#), INDIA. August 2002 - May 2006.
 - GPA: 9.43/10.0

PROFESSIONAL EXPERIENCE

- *Research Associate Professor*, Northwestern University, Mar. 2013 - present.
- *Research Assistant Professor*, Northwestern University, Sep. 2010 - Feb. 2013.
- *Postdoctoral Fellow*, Northwestern University, Sep. 2009 - Aug. 2010.
- *Research Assistant*, Iowa State University, Jan. 2007 - Aug. 2009.

RESEARCH GRANTS

1. **PI**, “The investigation of machine learning for material development”, *Toyota Motor Corporation*, \$200,000, 2017-2018.
2. **Senior Personnel**, “BD Spokes: SPOKE: MIDWEST: Collaborative: Integrative Materials Design (IMaD): Leverage, Innovate, & Disseminate”, *National Science Foundation (NSF)*, \$123,847 (Total \$989,700), 2017-2020. [[IIS-1636909](#)]
3. **PI**, “Social Media mining of caregiver experiences: Opportunity for preventing caregiver burnout”, *Northwestern Data Science Initiative*, \$25,000, 2017-2017.
4. **PI**, “Data-driven analytics for understanding processing-structure-property-performance relationships in steel alloys”, *Northwestern Data Science Initiative*, \$45,000, 2016-2017. [[DSI](#)]
5. **Co-PI**, “Scaling up the screening of molecular networks in the rational design of optically active materials”, *Northwestern Data Science Initiative*, \$9,000 (Total \$45,000), 2016-2017.
6. **PI**, “Analyzing caregiving experience on Twitter”, *Feinberg School of Medicine*, \$14,246, 2015-2016.

7. **Co-PI**, “SIMPLEX: Data-driven Discovery of Novel Thermoelectric Materials”, *Defense Advanced Research Projects Agency (DARPA)*, \$601,250 (Total \$1,559,999), 2015-2018. [[N66001-15-C-4036](#)]
8. **Co-PI**, “Scalable, In-situ Clustering and Data Analysis for Extreme Scale Scientific Computing”, *Department of Energy (DOE)*, \$1,219,899, 2015-2018. [[DE-SC0014330](#)]
9. **Co-PI**, “SHF:Medium:Collaborative Research: Scalable Algorithms for Spatio-temporal Data Analysis”, *National Science Foundation (NSF)*, \$709,342 (Total \$934,342), 2014-2018. [[CCF-1409601](#)]
10. **PI**, “Advanced Materials Center for Excellence: Center for Hierarchical Materials Design (CHiMaD)”, *National Institute of Standards and Technology (NIST)*, \$450,000 (Total \$25,000,000), 2014-2018. [[70NANB14H012](#)]
11. **Co-PI**, “EAGER: Scalable Big Data Analytics”, *National Science Foundation (NSF)*, \$300,000, 2013-2016. [[IIS-1343639](#)]
12. **Co-PI**, “MURI: MANAGING THE MOSAIC OF MICROSTRUCTURE: Image analysis, data structures, mathematical theory of microstructure, and hardware for the structure-property relationship”, *Air Force Office of Scientific Research (AFOSR), Department of Defense (DOD)*, \$750,000 (Total \$5,658,616), 2012-2017. [[FA9550-12-1-0458](#)]
13. **Northwestern Co-I**, “Scalable Data Management, Analysis, and Visualization (SDAV) Institute”, *Department of Energy (DOE)*, \$750,000 (Total \$25,000,000), 2012-2018. [[DE-SC0007456](#)]
14. **Senior Researcher**, “Expeditions in Computing: Understanding Climate Change: A Data Driven Approach”, *National Science Foundation (NSF)*, \$900,000 (Total \$10,000,000), 2010-2016. [[CCF-1029166](#)]
15. **Co-PI**, “EAGER: Discovering Knowledge from Scientific Research Networks”, *National Science Foundation (NSF)*, \$256,000, 2011-2014. [[ACI-1144061](#)]
16. **Research Participant**, “Scalable and Power Efficient Data Analytics for Hybrid Exascale Systems”, *Department of Energy (DOE)*, \$705,000, 2010-2014. [[DE-SC0005340](#)]

HONORS AND AWARDS

- Best Paper Finalist, HiPC 2017
- Northwestern Data Science Initiative Award, March 2017
- Northwestern Data Science Initiative Award, July 2016
- Best Paper Award at IEEE Cluster 2016
- Outstanding Paper Award awarded by Tata Consultancy Services for joint work published in IMMI, 2014
- Best Paper Award at KDD Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine), 2013
- Design Contest Winner at 25th International Conference on VLSI Design, 2012
- Research Excellence Award for Summer 2009 at ISU (to recognize up to 10% graduating students for outstanding research accomplishments)
- Graduate and Professional Student Senate Peer Research Award for Spring 2009 at ISU (based on research conducted at ISU)
- Best student paper award at International Conference on Information Technology, ICIT 2008
- Institute Silver Medal for obtaining the highest C.G.P.A. (9.43/10) amongst the students undergraduating in B.Tech. (Computer Science and Engineering) at IIT Roorkee, 2006
- Institute Silver Medal for Best Project in B. Tech. (Computer Science and Engineering) at IIT Roorkee, 2006

TEACHING

- Co-instructor: EECS 510: Social Media Mining, Spring 2017, Spring 2016, Spring 2015, Spring 2014, Northwestern University.
- Co-instructor: EECS 100: Electrons, Photons, and Bits: Adventures in Electrical and Computer Engineering, Spring 2017, Spring 2016, Spring 2015, Spring 2011, Northwestern University.

- Co-instructor: PSED 510-1: Predictive Science & Engineering Design Cluster Seminar, Fall 2017, Northwestern University.
- Co-instructor: MSIT 423: Data Mining and Business Intelligence, Spring 2011, Northwestern University.
- Teaching Assistant: ComS/BCB 567: Bioinformatics I - Fundamentals of Genome Informatics, Fall 2008, Iowa State University.
- Teaching Assistant: ComS 309: Software Engineering, Fall 2006, Iowa State University.

PUBLICATIONS *Book Chapters*

1. D. Han, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "A Fast DBSCAN Algorithm with Spark Implementation," in *Big Data in Engineering Applications*, S. S. R. et al., Ed. Springer Nature Singapore, 2019. To appear.
2. **A. Agrawal** and A. Choudhary, "Health Services Data: Big Data Analytics for Deriving Predictive Healthcare Insights," in *Data and Measures in Health Services Research*, B. Sobolev, A. Levy, and S. Goring, Eds. Springer US, 2016, pp. 1–17.
3. **A. Agrawal**, M. Patwary, W. Hendrix, W.-keng Liao, and A. Choudhary, "High performance big data clustering," in *Advances in Parallel Computing, Volume 23: Cloud Computing and Big Data*, L. Grandinetti, Ed. IOS Press, 2013, pp. 192–211.
4. **A. Agrawal**, A. Choudhary, and X. Huang, "Sequence-Specific Sequence Comparison Using Pairwise Statistical Significance," in *Software Tools and Algorithms for Biological Systems*, vol. 696, H. R. Arabnia, Ed. Springer, 2011, pp. 297–306.

Journal Publications

5. A. Furmanchuk, J. E. Saal, J. W. Doak, G. B. Olson, A. Choudhary, and **A. Agrawal**, "Prediction of seebeck coefficient for compounds without restriction to fixed stoichiometry: A machine learning approach," *Journal of Computational Chemistry*, vol. 39, no. 4, pp. 191–202, 2018.
6. M. K. Danilovich, J. Tsay, R. Al-Bahrani, A. Choudhary, and **A. Agrawal**, "#Alzheimer's and Dementia: Expressions of Memory Loss on Twitter," *Topics in Geriatric Rehabilitation*, vol. 34, pp. 48–53, 2018.
7. K. Gopalakrishnan, H. Gholami, A. Vidyadharan, A. Choudhary, and **A. Agrawal**, "Crack Damage Detection in Unmanned Aerial Vehicle Images of Civil Infrastructure Using Pre-trained Deep Learning Model," *International Journal for Traffic and Transport Engineering*, vol. 8, p. 1, 2018.
8. A. Paul, P. Acar, R. Liu, W.-keng Liao, A. Choudhary, V. Sundararaghavan, and **A. Agrawal**, "Data Sampling Schemes for Microstructure Design with Vibrational Tuning Constraints," *American Institute of Aeronautics and Astronautics (AIAA) Journal*, vol. 56, no. 3, pp. 1239–1250, 2018.
9. R. Al-Bahrani, **A. Agrawal**, and A. Choudhary, "Survivability prediction of colon cancer patients using neural networks," *Health Informatics Journal*, p. 1460458217720395, 2017.
10. Y. Cheng, **A. Agrawal**, H. Liu, and A. Choudhary, "Legislative Prediction with Dual Uncertainty Minimization from Heterogeneous Information," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 10, no. 2, pp. 110–120, 2017.
11. K. Gopalakrishnan, A. Choudhary, and **A. Agrawal**, "Big Data in Building Information Modeling Research: Survey and Exploratory Text Mining," *MOJ Civil Eng*, vol. 3, no. 6, p. 00087, 2017.
12. K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and **A. Agrawal**, "Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction and Building Materials*, vol. 157, pp. 322–330, 2017.

13. A. G. Gagorik, B. Savoie, N. Jackson, **A. Agrawal**, A. Choudhary, M. A. Ratner, G. C. Schatz, and K. L. Kohlstedt, "Improved Scaling of Molecular Network Calculations: The Emergence of Molecular Domains," *The Journal of Physical Chemistry Letters*, vol. 8, no. 2, pp. 415–421, 2017.
14. R. Liu, Y. C. Yabansu, Z. Yang, A. N. Choudhary, S. R. Kalidindi, and **A. Agrawal**, "Context Aware Machine Learning Approaches for Modeling Elastic Localization in Three-Dimensional Composite Microstructures," *Integrating Materials and Manufacturing Innovation*, pp. 1–12, 2017.
15. L. Ward, R. Liu, A. Krishna, V. I. Hegde, **A. Agrawal**, A. Choudhary, and C. Wolverton, "Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations," *Physical Review B*, vol. 96, no. 2, p. 024104, 2017.
16. Y. Xie, Z. Chen, D. Palsetia, G. Trajcevski, **A. Agrawal**, and A. Choudhary, "Silverback+: Scalable Association Mining Via Fast List Intersection For Columnar Social Data," *Knowledge and Information Systems (KAIS)*, vol. 50, no. 3, pp. 969–997, 2017.
17. **A. Agrawal** and A. Choudhary, "Perspective: Materials informatics and big data: Realization of the 'fourth paradigm' of science in materials science," *APL Materials*, vol. 4, no. 053208, pp. 1–10, 2016.
18. A. Furmanchuk, **A. Agrawal**, and A. Choudhary, "Predictive analytics for crystalline materials: Bulk modulus," *RSC Advances*, vol. 6, no. 97, pp. 95246–95251, 2016.
19. E. Rangel, W. Hendrix, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "AGORAS: A Fast Algorithm for Estimating Medoids in Large Datasets," *Procedia Computer Science*, vol. 80, pp. 1159–1169, 2016.
20. L. Ward, **A. Agrawal**, A. Choudhary, and C. Wolverton, "A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials," *npj Computational Materials*, vol. 2, no. 16028, 2016.
21. R. Liu, A. Kumar, Z. Chen, **A. Agrawal**, V. Sundararaghavan, and A. Choudhary, "A Predictive Machine Learning Approach for Microstructure Optimization and Materials Design," *Nature Scientific Reports*, vol. 5, no. 11551, 2015.
22. R. Liu, Y. C. Yabansu, **A. Agrawal**, S. R. Kalidindi, and A. N. Choudhary, "Machine learning approaches for elastic localization linkages in high-contrast composite materials," *Integrating Materials and Manufacturing Innovation*, vol. 4, no. 13, pp. 1–17, 2015.
23. **A. Agrawal**, P. D. Deshpande, A. Cecen, G. P. Basavarsu, A. N. Choudhary, and S. R. Kalidindi, "Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters," *Integrating Materials and Manufacturing Innovation*, vol. 3, no. 8, pp. 1–19, 2014.
24. A. R. Ganguly, E. Kodra, **A. Agrawal**, A. Banerjee, S. Boriah, S. Chatterjee, S. Chatterjee, A. Choudhary, D. Das, J. Faghmous, P. Ganguli, S. Ghosh, K. Hayhoe, C. Hays, W. Hendrix, Q. Fu, J. Kawale, D. Kumar, V. Kumar, W.-keng Liao, S. Liess, R. Mawalagedara, V. Mithal, R. Oglesby, K. Salvi, P. K. Snyder, K. Steinhäuser, D. Wang, and D. Wuebbles, "Toward enhanced understanding and projections of climate extremes using physics-guided data mining techniques," *Nonlinear Processes in Geophysics*, vol. 21, pp. 777–795, 2014.
25. B. Meredig, **A. Agrawal**, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary, and C. Wolverton, "Combinatorial screening for new materials in unconstrained composition space with machine learning," *Physical Review B*, vol. 89, no. 094104, pp. 1–7, 2014. **BM and AA are co-first authors.**
26. D. Palsetia, M. M. A. Patwary, **A. Agrawal**, and A. Choudhary, "Excavating Social Circles via User-Interests," *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 1–12, 2014.

27. S. W. Son, Z. Chen, W. Hendrix, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Data Compression for the Exascale Computing Era - Survey," *Supercomputing Frontiers and Innovations*, vol. 1, no. 2, pp. 76–88, 2014.
28. K. Gopalakrishnan, **A. Agrawal**, H. Ceylan, S. Kim, and A. Choudhary, "Knowledge discovery and data mining in pavement inverse analysis," *Transport*, vol. 28, no. 1, pp. 1–10, 2013.
29. J. S. Mathias, **A. Agrawal**, J. Feinglass, A. J. Cooper, D. W. Baker, and A. Choudhary, "Development of a 5 year life expectancy index in older adults using predictive mining of electronic health record data," *Journal of the American Medical Informatics Association*, vol. 20, pp. e118–e124, 2013. **JSM and AA are co-first authors.**
30. Y. Xie, Z. Chen, K. Zhang, Y. Cheng, D. K. Honbo, **A. Agrawal**, and A. Choudhary, "MuSES: a Multilingual Sentiment Elicitation System for Social Media Data," *IEEE Intelligent Systems*, vol. 99, pp. 1541–1672, 2013.
31. **A. Agrawal**, S. Misra, R. Narayanan, L. Polepeddi, and A. Choudhary, "Lung cancer survival prediction using ensemble data mining on SEER data," *Scientific Programming*, vol. 20, no. 1, pp. 29–42, 2012.
32. Y. Zhang, S. Misra, **A. Agrawal**, M. M. A. Patwary, W.-keng Liao, Z. Qin, and A. Choudhary, "Accelerating pairwise statistical significance estimation for local alignment by harvesting GPU's power," *BMC Bioinformatics*, vol. 13, no. Suppl 5, p. S3, 2012.
33. Y. Zhang, M. M. A. Patwary, S. Misra, **A. Agrawal**, W.-keng Liao, Z. Qin, and A. Choudhary, "Par-psse: Software for pairwise statistical significance estimation in parallel for local sequence alignment," *International Journal of Digital Content Technology and its Applications (JDCTA)*, vol. 6, no. 5, pp. 200–208, 2012.
34. Y. Zhang, F. Zhou, J. Gou, H. Xiao, Z. Qin, and **A. Agrawal**, "Accelerating Pairwise Statistical Significance Estimation Using NUMA Machine," *Journal of Computational Information Systems (JCIS)*, vol. 8, no. 9, pp. 3887–3894, 2012.
35. **A. Agrawal** and A. Choudhary, "Association Rule Mining Based HotSpot Analysis on SEER Lung Cancer Data," *International Journal of Knowledge Discovery in Bioinformatics (IJKDB)*, vol. 2, no. 2, pp. 34–54, 2011.
36. **A. Agrawal** and X. Huang, "Pairwise Statistical Significance of Local Sequence Alignment Using Sequence-Specific and Position-Specific Substitution Matrices," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 8, no. 1, pp. 194–205, 2011.
37. **A. Agrawal**, S. Misra, D. Honbo, and A. Choudhary, "Parallel pairwise statistical significance estimation of local sequence alignment using Message Passing Interface library," *Concurrency and Computation: Practice and Experience*, vol. 23, no. 17, pp. 2269–2279, 2011.
38. S. Misra, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Anatomy of a hash-based long read sequence mapping algorithm for next generation DNA sequencing," *Bioinformatics*, vol. 27, no. 2, pp. 189–195, 2011.
39. **A. Agrawal**, A. Mittal, R. Jain, and R. Takkar, "Fuzzy-adaptive-thresholding-based exon prediction," *International Journal of Computational Biology and Drug Design*, vol. 3, no. 4, pp. 311–333, 2010.
40. **A. Agrawal** and X. Huang, "Pairwise Statistical Significance of Local Sequence Alignment Using Multiple Parameter Sets and Empirical Justification of Parameter Set Change Penalty," *BMC Bioinformatics*, vol. 10, no. Suppl 3, p. S1, 2009.
41. **A. Agrawal** and X. Huang, "PSIBLAST_PairwiseStatSig: Reordering PSI-BLAST hits using pairwise statistical significance," *Bioinformatics*, vol. 25, no. 8, pp. 1082–1083, 2009.

42. **A. Agrawal**, V. P. Brendel, and X. Huang, "Pairwise Statistical Significance and Empirical Determination of Effective Gap Opening Penalties for Protein Local Sequence Alignment," *International Journal of Computational Biology and Drug Design*, vol. 1, no. 4, pp. 347–367, 2008.
43. **A. Agrawal** and A. Mittal, "A dynamic time-lagged correlation based method to learn multi-time delay gene networks," *World Academy of Science, Engineering and Technology*, vol. 9, pp. 723–730, 2007.
44. **A. Agrawal**, A. Mittal, and S. Gupta, "A Restrictive Mining Algorithm to Learn Multi Time Delay Gene Networks," *Bioinformatics Trends - A Journal of Bioinformatics and Its Applications*, vol. 2, no. 3, 4, pp. 125–140, 2007.

Peer-Reviewed Conference Publications

45. R. Al-Bahrani, M. K. Danilovich, W.-keng Liao, A. Choudhary, and **A. Agrawal**, "Towards Identifying Informal Caregivers of Alzheimer's and Dementia Patients in Social Media," in *Proceedings of the Fifth International Conference on Healthcare Informatics (ICHI)*, 2017, pp. 324–324.
46. K. Lee, **A. Agrawal**, and A. Choudhary, "Forecasting Influenza Levels using Real-Time Social Media Streams," in *Proceedings of the Fifth International Conference on Healthcare Informatics (ICHI)*, 2017, pp. 409–414.
47. K. Lee, S. A. Hasan, O. Farri, A. Choudhary, and **A. Agrawal**, "Medical Concept Normalization for Online User-Generated Texts," in *Proceedings of the Fifth International Conference on Healthcare Informatics (ICHI)*, 2017, pp. 462–469.
48. S. Lee, D. Jha, **A. Agrawal**, A. Choudhary, and W.-keng Liao, "Parallel Deep Convolutional Neural Network Training by Exploiting the Overlapping of Computation and Communication," in *Proceedings of 24th IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC'17)*, 2017, pp. 183–192.
49. E. Rangel, N. Frontiere, S. Habib, K. Heitmann, W.-keng Liao, **A. Agrawal**, and A. Choudhary, "Building Halo Merger Trees from the Q Continuum Simulation," in *Proceedings of 24th IEEE International Conference on High Performance Computing, Data, and Analytics (HiPC'17)*, 2017, pp. 398–407.
50. Y. Xie, Z. Chen, **A. Agrawal**, and A. Choudhary, "Distinguish polarity in bag-of-words visualization," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)*, 2017, pp. 3344–3350.
51. **A. Agrawal** and A. Choudhary, "A Fatigue Strength Predictor for Steels Using Ensemble Data Mining," in *Proceedings of 25th ACM International Conference on Information and Knowledge Management (CIKM) (Demo)*, 2016, pp. 2497–2500.
52. **A. Agrawal**, J. Mathias, D. Baker, and A. Choudhary, "Identifying HotSpots in Five Year Survival Electronic Health Records of Older Adults," in *Proceedings of 6th IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS)*, 2016, pp. 1–6.
53. **A. Agrawal**, J. Mathias, D. Baker, and A. Choudhary, "Five Year Life Expectancy Calculator for Older Adults," in *Proceedings of IEEE International Conference on Data Mining (ICDM) (Demo)*, 2016, pp. 1280–1283.
54. **A. Agrawal**, B. Meredig, C. Wolverton, and A. Choudhary, "A Formation Energy Predictor for Crystalline Materials Using Ensemble Data Mining," in *Proceedings of IEEE International Conference on Data Mining (ICDM) (Demo)*, 2016, pp. 1276–1279.
55. R. Al-Bahrani, M. K. Danilovich, **A. Agrawal**, and A. Choudhary, "Towards Informal Caregiver Identification in Social Media," in *Proceedings of IEEE International Conference on Healthcare Informatics (ICHI) (Poster)*, 2016, pp. 414–414.

56. Q. Kang, W.-keng Liao, **A. Agrawal**, and A. Choudhary, "A Filtering-based Clustering Algorithm for Improving Spatio-temporal Kriging Interpolation Accuracy," in Proceedings of 25th ACM International Conference on Information and Knowledge Management (CIKM), 2016, pp. 2209–2214.
57. D. Palsetia, W. Hendrix, S. Lee, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Parallel Community Detection Algorithm Using a Data Partitioning Strategy with Pairwise Subdomain Duplication," in High Performance Computing, 31st International Conference, ISC High Performance 2016, Frankfurt, Germany, June 19–23, 2016, Proceedings, 2016, pp. 98–115.
58. E. Rangel, N. Li, S. Habib, T. Peterka, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Parallel DTFE Surface Density Field Reconstruction," in Proceedings of IEEE Cluster, 2016, pp. 30–39. Won the best paper award.
59. Z. Yuan, W. Hendrix, S. W. Son, C. Federrath, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Parallel Implementation of Lossy Data Compression for Temporal Data Sets," in Proceedings of 23rd Annual International Conference on High Performance Computing, Data, and Analytics (HiPC'16), 2016, pp. 62–71.
60. Y. Cheng, **A. Agrawal**, H. Liu, and A. Choudhary, "Legislative prediction with dual uncertainty minimization from heterogeneous information," in Proceedings of the 15th SIAM International Conference on Data Mining (SDM), 2015, pp. 361–369.
61. C. Jin, Z. Chen, W. Hendrix, **A. Agrawal**, and A. Choudhary, "Incremental, Distributed Single-Linkage Hierarchical Clustering Algorithm Using MapReduce," in Proceedings of the 23rd High Performance Computing Symposium (HPC), 2015, pp. 83–92.
62. C. Jin, Q. Fu, H. Wang, W. Hendrix, Z. Chen, **A. Agrawal**, A. Banerjee, and A. Choudhary, "Running MAP Inference on Million Node Graphical Models: A High Performance Computing Perspective," in Proceedings of the 15th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), 2015, pp. 565–575.
63. C. Jin, R. Liu, Z. Chen, W. Hendrix, **A. Agrawal**, and A. Choudhary, "A Scalable Hierarchical Clustering Algorithm Using Spark," in Proceedings of IEEE International Conference on Big Data Computing Service and Applications (BigDataService), 2015, pp. 418–426.
64. K. Lee, **A. Agrawal**, and A. Choudhary, "Mining Social Media Streams to Improve Public Health Allergy Surveillance," in Proceedings of IEEE/ACM International Conference on Social Networks Analysis and Mining (ASONAM), 2015, pp. 815–822.
65. R. Liu, **A. Agrawal**, Z. Chen, W.-keng Liao, and A. Choudhary, "Pruned Search: A Machine Learning Based Meta-Heuristic Approach for Constrained Continuous Optimization," in Proceedings of 8th IEEE International Conference on Contemporary Computing (IC3), 2015, pp. 13–18.
66. V. Rastogi and **A. Agrawal**, "All Your Google and Facebook Logins are Belong to Us: A Case for Single Sign-off," in Proceedings of 8th IEEE International Conference on Contemporary Computing (IC3), 2015, pp. 416–421.
67. Y. Xie, P. Daga, Y. Cheng, K. Zhang, **A. Agrawal**, and A. Choudhary, "Reducing infrequent-token perplexity via variational corpora," in Proceedings of the 53rd Annual Meeting of the Association of Computational Linguistics (ACL) and the 7th International Joint Conference on Natural Language Processing, 2015, pp. 609–615.
68. Y. Cheng, **A. Agrawal**, H. Liu, and A. Choudhary, "Social Role Identification via Dual Uncertainty Minimization Regularization," in Proceedings of International Conference on Data Mining (ICDM), 2014, pp. 767–772.
69. Z. Chen, S. W. Son, W. Hendrix, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "NUMARCK: Machine Learning Algorithm for Resiliency and Checkpointing," in Proceedings of 26th International Conference on High Performance Computing, Networking, Storage and Analysis (Supercomputing, SC'14), 2014, pp. 733–744.

70. D. Palsetia, M. Patwary, W. Hendrix, **A. Agrawal**, and A. Choudhary, "Clique Guided Community Detection," in Proceedings of 2014 IEEE International Conference on Big Data (BigData), 2014, pp. 500–509.
71. Y. Xie, Z. Chen, D. Palsetia, **A. Agrawal**, and A. Choudhary, "Indexing Bipartite Memberships in Web Graphs," in Proceedings of IEEE/ACM International Conference on Social Networks Analysis and Mining (ASONAM) 2014, 2014, pp. 166–173.
72. Y. Xie, D. Palsetia, G. Trajcevski, **A. Agrawal**, and A. Choudhary, "Silverback: Scalable Association Mining For Temporal Data in Columnar Probabilistic Databases," in Proceedings of 30th IEEE International Conference on Data Engineering (ICDE), Industrial and Applications Track, 2014, pp. 1072–1083.
73. Y. Cheng, Z. Chen, L. Liu, J. Wang, **A. Agrawal**, and A. Choudhary, "Feedback-Driven Multiclass Active Learning for Data Streams," in Proceedings of 22th ACM International Conference on Information and Knowledge Management (CIKM 2013), San Francisco, USA, Oct. 2013, 2013, pp. 1311–1320.
74. Y. Cheng, Z. Chen, J. Wang, **A. Agrawal**, and A. Choudhary, "Bootstrapping Active Name Disambiguation with Crowdsourcing," in Proceedings of 22th ACM International Conference on Information and Knowledge Management (CIKM 2013), San Francisco, USA, Oct. 2013 (Poster paper), 2013, pp. 1213–1216.
75. Z. Chen, Y. Xie, Y. Cheng, K. Zhang, **A. Agrawal**, W.-keng Liao, N. Samatova, and A. Choudhary, "Forecast Oriented Classification of Spatio-Temporal Extreme Events," in Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013, pp. 2952–2954.
76. Y. Cheng, Y. Xie, Z. Chen, **A. Agrawal**, A. Choudhary, and S. Guo, "JobMiner: A Real-time System for Mining Job-related Patterns from Social Media," in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD) (Demo paper), 2013, pp. 1450–1453.
77. P. D. Deshpande, B. P. Gautham, A. Cecen, S. Kalidindi, **A. Agrawal**, and A. Choudhary, "Application of Statistical and Machine Learning Techniques for Correlating Properties to Composition and Manufacturing Processes of Steels," in 2nd World Congress on Integrated Computational Materials Engineering, July 7-11, 2013, Salt Lake City, Utah, 2013, pp. 155–160.
78. W. Hendrix, D. Palsetia, M. M. A. Patwary, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "A Scalable Algorithm for Single-Linkage Hierarchical Clustering on Distributed Memory Architectures," in Proceedings of 3rd IEEE Symposium on Large-Scale Data Analysis and Visualization (LDAV 2013), Atlanta GA, USA, Oct. 2013, 2013, pp. 7–13.
79. K. Lee, **A. Agrawal**, and A. Choudhary, "Real-Time Disease Surveillance using Twitter Data: Demonstration on Flu and Cancer," in Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD) (Demo paper), 2013, pp. 1474–1477.
80. L. Liu, J. Tang, Y. Cheng, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Mining Diabetes Complication and Treatment Patterns for Clinical Decision Support," in Proceedings of 22th ACM International Conference on Information and Knowledge Management (CIKM 2013), San Francisco, USA, Oct. 2013, 2013, pp. 279–288.
81. M. Patwary, D. Palsetia, **A. Agrawal**, W.-keng Liao, F. Manne, and A. Choudhary, "Scalable Parallel OPTICS Data Clustering Using Graph Algorithmic Techniques," in Proceedings of 25th International Conference on High Performance Computing, Networking, Storage and Analysis (Supercomputing, SC'13), 2013, pp. 1–12. Article No. 49.
82. Y. Xie, Z. Chen, **A. Agrawal**, A. Choudhary, and L. Liu, "Random Walk-based Graphical Sampling in Unbalanced Heterogeneous Bipartite Social Graphs," in Proceedings of 22th

- ACM International Conference on Information and Knowledge Management (CIKM 2013), San Francisco, USA, Oct. 2013 (Poster paper), 2013, pp. 1473–1476.
83. Y. Xie, Z. Chen, **A. Agrawal**, W.-keng Liao, and A. Choudhary, “Caranx: Scalable Social Image Index Using Phylogenetic Tree of Hashtags,” in Proceedings of 25th International Conference on High Performance Computing, Networking, Storage and Analysis (Supercomputing, SC’13) (Poster paper), 2013, pp. 1–2.
 84. Y. Xie, Z. Chen, Y. Cheng, K. Zhang, **A. Agrawal**, W.-keng Liao, and A. Choudhary, “Detecting and Tracking Disease Outbreaks by Mining Social Media Data,” in Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI), 2013, pp. 2958–2960.
 85. Y. Xie, Z. Chen, K. Zhang, M. M. A. Patwary, Y. Cheng, H. Liu, **A. Agrawal**, and A. Choudhary, “Graphical Modeling of Macro Behavioral Targeting in Social Networks,” in Proceedings of the 13th SIAM International Conference on Data Mining (SDM), 2013, pp. 740–748.
 86. Y. Xie, Z. Chen, K. Zhang, Y. Cheng, C. Jin, **A. Agrawal**, and A. Choudhary, “Elver: Recommending Facebook Pages in Cold Start Situation Without Content Features,” in Proceedings of IEEE International Conference on Big Data (IEEE BigData 2013), 2013, pp. 475–479.
 87. K. Zhang, D. Downey, Z. Chen, Y. Xie, Y. Cheng, **A. Agrawal**, W.-keng Liao, and A. Choudhary, “A Probabilistic Graphical Model for Brand Reputation Assessment in Social Networks,” in Proceedings of IEEE/ACM International Conference on Social Networks Analysis and Mining (ASONAM), Niagara Falls, Canada, Aug. 2013, 2013, pp. 223–230.
 88. Y. Cheng, K. Zhang, Y. Xie, **A. Agrawal**, and A. Choudhary, “On active learning in hierarchical classification,” in Proceedings of the 21st ACM International Conference on Information and Knowledge Management (CIKM) (Poster paper), 2012, pp. 2467–2470.
 89. W. Hendrix, M. M. A. Patwary, **A. Agrawal**, W.-keng Liao, and A. Choudhary, “Parallel Hierarchical Clustering on Shared Memory Platforms,” in Proceedings of the 19th International Conference on High Performance Computing (HiPC), 2012, pp. 1–9.
 90. M. Patwary, D. Palsetia, **A. Agrawal**, W.-keng Liao, F. Manne, and A. Choudhary, “A new scalable parallel DBSCAN algorithm using the disjoint-set data structure,” in ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC), 2012, pp. 1–11. Article No. 62.
 91. Y. Xie, D. Honbo, A. Choudhary, K. Zhang, Y. Cheng, and **A. Agrawal**, “VOXSUP: a social engagement framework,” in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD) (Demo paper), 2012, pp. 1556–1559.
 92. K. Zhang, Y. Xie, Y. Cheng, D. Honbo, D. Downey, **A. Agrawal**, W.-keng Liao, and A. Choudhary, “Sentiment identification by incorporating syntax, semantics and context information,” in Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval (Poster paper), 2012, pp. 1143–1144.
 93. **A. Agrawal**, S. Misra, A. Choudhary, and K. Bilimoria, “Risk prediction for post-operative adverse outcomes in colorectal cancer surgery,” in Proc. of IEEE 1st International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) (Poster paper), 2011, pp. 232–232.
 94. **A. Agrawal**, S. Misra, R. Narayanan, L. Polepeddi, and A. Choudhary, “A lung cancer mortality risk calculator based on SEER data,” in Proc. of IEEE 1st International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) (Poster paper), 2011, pp. 233–233.
 95. Y. Zhang, S. Misra, D. Honbo, **A. Agrawal**, W.-keng Liao, and A. Choudhary, “Efficient pairwise statistical significance estimation for local sequence alignment using GPU,” in IEEE 1st International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), 2011, pp. 226–231.

96. **A. Agrawal**, A. Choudhary, and X. Huang, “Derived Distribution Points Heuristic for Fast Pairwise Statistical Significance Estimation,” in Proceedings of the First ACM International Conference on Bioinformatics and Computational Biology (BCB), 2010, pp. 312–321.
97. **A. Agrawal**, A. Choudhary, and X. Huang, “Non-conservative Pairwise Statistical Significance of Local Sequence Alignment Using Position-Specific Substitution Matrices,” in Proceedings of BIOCOMP, 2010, pp. 262–268.
98. D. Honbo, **A. Agrawal**, and A. Choudhary, “Efficient pairwise statistical significance estimation using FPGAs,” in Proceedings of BIOCOMP, 2010, pp. 571–577.
99. **A. Agrawal**, V. Brendel, and X. Huang, “Pairwise Statistical Significance Versus Database Statistical Significance for Local Alignment of Protein Sequences,” in Bioinformatics Research and Applications, 2008, vol. 4983, pp. 50–61.
100. **A. Agrawal**, A. Ghosh, and X. Huang, “Estimating Pairwise Statistical Significance of Protein Local Alignments Using a Clustering-Classification Approach Based on Amino Acid Composition,” in Bioinformatics Research and Applications, 2008, vol. 4983, pp. 62–73.
101. **A. Agrawal** and X. Huang, “Pairwise Statistical Significance of Local Sequence Alignment Using Substitution Matrices with Sequence-Pair-Specific Distance,” in Proc. of International Conference on Information Technology, ICIT, 2008, pp. 94–99. **Won the best student paper award.**
102. **A. Agrawal** and X. Huang, “Pairwise DNA Alignment with Sequence Specific Transition-Transversion Ratio Using Multiple Parameter Sets,” in Proc. of International Conference on Information Technology, ICIT, 2008, pp. 89–93.
103. **A. Agrawal** and X. Huang, “Conservative, Non-conservative and Average Pairwise Statistical Significance of Local Sequence Alignment,” in Proc. of IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2008, pp. 433–436.
104. **A. Agrawal** and X. Huang, “DNAlignTT: Pairwise DNA Alignment with Sequence Specific Transition-Transversion Ratio,” in Proc. of IEEE International Conference on Electro/Information Technology (EIT), 2008, pp. 453–455.
105. **A. Agrawal** and S. K. Khaitan, “A new heuristic for multiple sequence alignment,” in Proc. of IEEE International Conference on Electro/Information Technology (EIT) (Poster paper), 2008, pp. 215–217.
106. **A. Agrawal**, A. Mittal, R. Jain, and R. Takkar, “An adaptive fuzzy thresholding algorithm for exon prediction,” in Proc. of IEEE International Conference on Electro/Information Technology (EIT) (Poster paper), 2008, pp. 211–214.
107. **A. Agrawal** and A. Mittal, “Identifying temporal gene networks using signal processing metrics on time-series gene expression data,” in Proc. of 3rd International Conference on Intelligent Sensing and Information Processing (ICISIP), 2005, pp. 86–92.
108. A. Mittal, S. Gupta, **A. Agrawal**, and L. F. Cheong, “Camera Motion Characterization of Establishment Shots and Video Signatures,” in Proc. of 12th International Conference on Advanced Computing and Communications (ADCOM), 2004, pp. 663–668.
109. **A. Agrawal**, A. Mittal, and S. Gupta, “Identifying Temporal Gene Networks by Mining Gene Expression Data,” in Proc. of 12th International Conference on Advanced Computing and Communications (ADCOM), 2004, pp. 194–200.

Peer-Reviewed Workshop Publications

110. R. Al-Bahrani, M. K. Danilovich, W.-keng Liao, A. Choudhary, and **A. Agrawal**, “Analyzing Informal Caregiving Expression in Social Media,” in IEEE ICDM Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction (SENTIRE), 2017, pp. 342–349.

111. Q. Kang, W.-keng Liao, **A. Agrawal**, and A. Choudhary, "A Hybrid Training Algorithm for Recurrent Neural Network Using Particle Swarm Optimization-based Preprocessing and Temporal Error Aggregation," in IEEE ICDM Workshop on Optimization Based Techniques for Emerging Data Mining Problems (OEDM), 2017, pp. 812–817.
112. D. Han, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "A novel scalable DBSCAN algorithm with Spark," in Proceedings of 5th IEEE IPDPS Workshop on Parallel and Distributed Computing for Large Scale Machine Learning and Big Data Analytics (ParLearning), 2016, pp. 1393–1402.
113. A. Krishna, **A. Agrawal**, and A. Choudhary, "Predicting the Outcome of Startups: Less Failure, More Success," in Proceedings of IEEE ICDM Workshop on Data Market for Co-evolution of Sciences and Business (MoDAT), 2016, pp. 798–805.
114. R. Liu, **A. Agrawal**, W.-keng Liao, M. D. Graef, and A. Choudhary, "Materials Discovery: Understanding Polycrystals from Large-Scale Electron Patterns," in Proceedings of IEEE Big-Data Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH), 2016, pp. 2261–2269.
115. S. Lee, W.-keng Liao, **A. Agrawal**, N. Hardavellas, and A. Choudhary, "Evaluation of K-Means Data Clustering Algorithm on Intel Xeon Phi," in Proceedings of IEEE BigData Workshop on Advances in Software and Hardware for Big Data to Knowledge Discovery (ASH), 2016, pp. 2251–2260.
116. R. Liu, D. Palsetia, A. Paul, R. Al-Bahrani, D. Jha, W.-keng Liao, **A. Agrawal**, and A. Choudhary, "PinterNet: A Thematic Label Curation Tool for Large Image Datasets," in Proceedings of IEEE BigData Workshop on Open Science in Big Data (OSBD), 2016, pp. 2353–2362.
117. R. Liu, L. Ward, C. Wolverton, **A. Agrawal**, W.-K. Liao, and A. Choudhary, "Deep Learning for Chemical Compound Stability Prediction," in Proceedings of ACM SIGKDD Workshop on Large-scale Deep Learning for Data Mining (DL-KDD), 2016, pp. 1–7.
118. A. Paul, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "AnonyMine: Mining anonymous social media posts using psycho-lingual and crowd-sourced dictionaries," in Proceedings of ACM SIGKDD Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM), 2016.
119. S. Gupta, D. Palsetia, M. M. A. Patwary, **A. Agrawal**, and A. Choudhary, "A New Parallel Algorithm for Two-Pass Connected Component Labeling," in Proceedings of IEEE IPDPS Workshop on Multithreaded Architectures and Applications (MTAAP), 2014, pp. 1355–1362.
120. R. Liu, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Enhancing Financial Decision-Making Using Social Behavior Modeling," in Proceedings of 8th KDD Workshop on Social Network Mining and Analysis for Business, Consumer and Social Insights (SNAKDD), 2014, pp. 13:1–13:5. Article No. 13.
121. R. Liu, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Search Space Preprocessing in Solving Complex Optimization Problems," in Proceedings of IEEE BigData Workshop on Complexity for Big Data (C4BD), 2014, pp. 1–5.
122. Y. Xie, Y. Cheng, **A. Agrawal**, and A. Choudhary, "Estimating online user location distribution without GPS location," in Proceedings of ICDM Workshop on Connecting Online and Offline Social Network Analysis (COOL-SNA), 2014, pp. 936–943.
123. **A. Agrawal**, R. Al-Bahrani, R. Merkow, K. Bilimoria, and A. Choudhary, "Colon Surgery Outcome Prediction Using ACS NSQIP Data," in Proceedings of the KDD Workshop on Data Mining for Healthcare (DMH), 2013, pp. 1–6.
124. **A. Agrawal**, R. Al-Bahrani, J. Raman, M. J. Russo, and A. Choudhary, "Lung Transplant Outcome Prediction using UNOS Data," in Proceedings of the IEEE Big Data Workshop on Bioinformatics and Health Informatics (BHI), 2013, pp. 1–8.

125. **A. Agrawal** and A. Choudhary, "An Analysis of Variation in Hospital Billing Using Medicare Data," in Proceedings of the KDD Workshop on Data Mining for Healthcare (DMH), 2013, pp. 1–6.
126. **A. Agrawal**, J. Raman, M. J. Russo, and A. Choudhary, "Heart Transplant Outcome Prediction using UNOS Data," in Proceedings of the KDD Workshop on Data Mining for Healthcare (DMH), 2013, pp. 1–6.
127. R. Al-Bahrani, **A. Agrawal**, and A. Choudhary, "Colon cancer survival prediction using ensemble data mining on SEER data," in Proceedings of the IEEE Big Data Workshop on Bioinformatics and Health Informatics (BHI), 2013, pp. 9–16.
128. C. Jin, Q. Fu, H. Wang, **A. Agrawal**, W. Hendrix, W.-keng Liao, M. M. A. Patwary, A. Banerjee, and A. Choudhary, "Solving Combinatorial Optimization Problems using Relaxed Linear Programming: A High Performance Computing Perspective," in Proceedings of the KDD Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine), 2013, pp. 39–46. **Won the best paper award.**
129. C. Jin, M. M. A. Patwary, **A. Agrawal**, W. Hendrix, W.-keng Liao, and A. Choudhary, "DiSC: A Distributed Single-Linkage Hierarchical Clustering Algorithm using MapReduce," in Proceedings of the 4th International SC Workshop on Data Intensive Computing in the Clouds (DataCloud) 2013, 2013, pp. 1–6.
130. K. Lee, **A. Agrawal**, and A. Choudhary, "Real-Time Digital Flu Surveillance using Twitter Data," in Proceedings of the SDM Workshop on Data Mining for Medicine and Healthcare (DMMH), 2013, pp. 19–27.
131. Y. Cheng, Y. Xie, K. Zhang, **A. Agrawal**, and A. Choudhary, "CluChunk: clustering large scale user-generated content incorporating chunklet information," in Proceedings of the KDD Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications (BigMine), 2012, pp. 12–19.
132. Y. Cheng, Y. Xie, K. Zhang, **A. Agrawal**, and A. Choudhary, "How Online Content is Received by Users in Social Media: A Case Study on Facebook. com Posts," in Proceedings of the KDD Workshop on Social Media Analytics (SOMA), 2012, pp. 1–8.
133. D. Palsetia, M. M. A. Patwary, K. Zhang, K. Lee, C. Moran, Y. Xie, D. Honbo, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "User-Interest based Community Extraction in Social Networks," in Proceedings of the KDD Workshop on Social Network Mining and Analysis (SNAKDD), 2012, pp. 1–4.
134. Y. Xie, Y. Cheng, D. Honbo, K. Zhang, **A. Agrawal**, A. Choudhary, Y. Gao, and J. Gou, "Probabilistic macro behavioral targeting," in Proceedings of the CIKM workshop on Data-driven user behavioral modelling and mining from social media (DUBMMSM), 2012, pp. 7–10.
135. Y. Xie, Y. Cheng, D. Honbo, K. Zhang, **A. Agrawal**, and A. Choudhary, "Crowdsourcing recommendations from social sentiment," in Proceedings of KDD Workshop on Issues of Sentiment Discovery and Opinion Mining (WISDOM), 2012, pp. 9:1–9:8.
136. **A. Agrawal** and A. Choudhary, "Identifying HotSpots in Lung Cancer Data Using Association Rule Mining," in 2nd IEEE ICDM Workshop on Biological Data Mining and its Applications in Healthcare (BioDM), 2011, pp. 995–1002.
137. **A. Agrawal**, S. Misra, R. Narayanan, L. Polepeddi, and A. Choudhary, "A lung cancer outcome calculator using ensemble data mining on SEER data," in Proceedings of the Tenth International Workshop on Data Mining in Bioinformatics (BIOKDD), New York, NY, USA, 2011, pp. 1–9.
138. Y. Cheng, K. Zhang, Y. Xie, **A. Agrawal**, W.-keng Liao, and A. Choudhary, "Learning to Group Web Text Incorporating Prior Information," in 6th IEEE ICDM Workshop on Optimization Based Techniques for Emerging Data Mining Problems, (OEDM), 2011, pp. 212–219.

139. W. Hendrix, I. Tetteh, **A. Agrawal**, F. Semazzi, W.-keng Liao, and A. Choudhary, "Community Dynamics and Analysis of Decadal Trends in Climate Data," in 3rd IEEE ICDM Workshop on Knowledge Discovery from Climate Data, (ClimKD), 2011, pp. 9–14.
140. K. Lee, D. Palsetia, R. Narayanan, M. Patwary, **A. Agrawal**, and A. Choudhary, "Twitter Trending Topic Classification," in 6th IEEE ICDM Workshop on Optimization Based Techniques for Emerging Data Mining Problems, (OEDM), 2011, pp. 251–258.
141. N. Nakka, **A. Agrawal**, and A. N. Choudhary, "Predicting Node Failure in High Performance Computing Systems from Failure and Usage Logs," in IEEE International Symposium on Parallel and Distributed Processing Symposium (IPDPS) Workshops, 2011, pp. 1557–1566.
142. L. Polepeddi, **A. Agrawal**, and A. Choudhary, "Poll: A Citation-Text-Based System for Identifying High-Impact Contributions of an Article," in IEEE ICDM Workshop on Data Mining in Networks, (DaMNet), 2011, pp. 965–968.
143. K. Zhang, Y. Cheng, Y. Xie, **A. Agrawal**, D. Palsetia, K. Lee, and A. Choudhary, "SES: Sentiment Elicitation System for Social Media Data," in IEEE ICDM Workshop on Sentiment Elicitation from Natural Text for Information Retrieval and Extraction, (SENTIRE), 2011, pp. 129–136.
144. Y. Zhang, M. Patwary, S. Misra, **A. Agrawal**, W.-keng Liao, and A. N. Choudhary, "Enhancing parallelism of pairwise statistical significance estimation for local sequence alignment," in HiPC Workshop on Hybrid Multi-core Computing (WHMC), 2011, pp. 1–8.
145. **A. Agrawal**, S. Misra, D. Honbo, and A. Choudhary, "MPIPairwiseStatSig: Parallel Pairwise Statistical Significance Estimation of Local Sequence Alignment," in Proceedings of the HPDC Workshop on Emerging Computational Methods for the Life Sciences (ECMLS), 2010, pp. 470–476.
146. **A. Agrawal** and X. Huang, "Pairwise Statistical Significance of Local Sequence Alignment Using Multiple Parameter Sets," in Proc. of ACM 2nd International Workshop on Data and Text Mining in Bioinformatics (DTMBIO), 2008, pp. 53–60.

SCIENTIFIC
PRESENTATIONS
AND TALKS

1. (Invited) Data-Driven Approaches for Steel Fatigue Strength Prediction , *TMS 2018: Fatigue in Materials: Fundamentals, Multiscale Modeling and Prevention - Data-driven Investigations of Fatigue*, March 12, 2018, Phoenix AZ, USA.
2. (Invited) Materials Informatics and Big Data: Realization of "Fourth Paradigm" of Science in Materials Science, *Materials Science Division Colloquium, Argonne National Lab*, February 15, 2018, Lemont IL, USA.
3. Deep Learning Models for Structure-Property Linkages in High Contrast Composites, *MURI Final Review*, December 18, 2017, Arlington VA, USA.
4. Learning Crystal Orientations of Polycrystalline Materials from Electron Backscatter Diffraction Experiments using Convolutional Neural Networks, *MURI Final Review*, December 18, 2017, Arlington VA, USA.
5. (Invited) Materials Informatics on Images: Structure Characterization, Crack Detection, Localization, and More, *Machine Learning Applied to Materials Imaging Workshop, Northwestern-Argonne Institute of Science and Engineering*, October 30, 2017, Evanston IL, USA.
6. (Invited) High Performance Data Mining: An Essential Paradigm for Interdisciplinary Big Data Analytics, *2017 IEEE Region 4 Workshop on Big Data*, October 25, 2017, Evanston IL, USA.
7. Materials Informatics and Big Data: Realization of "Fourth Paradigm" of Science in Materials Science, *MS&T 2017: Data and Tools for Materials Discovery and Design: Data Science Methods in Materials Discovery and Development*, October 11, 2017, Pittsburgh PA, USA.
8. Data-Driven Approaches for Predicting Thermoelectric Properties, *MS&T 2017: In-situ Characterization of Energy Materials*, October 10, 2017, Pittsburgh PA, USA.

9. Classification of Scientific Journal Articles to Support Automated Data Extraction and Curation, *MS&T 2017: Recent Advances in Computer-aided Materials Design: Emerging Approaches of Material Design*, October 10, 2017, Pittsburgh PA, USA.
10. Data-Driven Approaches for Predicting Fatigue Strength of Steels, *MS&T 2017: Shaping & Forming of Advanced High Strength Steels: Performance*, October 10, 2017, Pittsburgh PA, USA.
11. Materials Informatics and Big Data: Realization of 4th Paradigm of Science in Materials Science, *Materials Research and Data Science Conference*, September 25, 2017, Rockville MD, USA.
12. (Invited) Materials Informatics and Big Data: Realization of 4th Paradigm of Science in Materials Science, *MAPEX Symposium 2017, University of Bremen*, September 15, 2017, Bremen, Germany.
13. (Invited) Materials Informatics and Big Data: Realization of 4th Paradigm of Science in Materials Science, *Materials Genome Symposium, Chinese Materials Conference (CMC 2017), Chinese Materials Research Society (C-MRS)*, July 11, 2017, Yinchuan, Ningxia, China.
14. Materials Informatics and Big Data: Realization of 4th Paradigm of Science in Materials Science, *4th World Congress on Integrated Computational Materials Engineering (ICME 2017), ICME Success Stories and Applications*, May 24, 2017, Ypsilanti MI, USA.
15. (Invited) Materials Informatics and Big Data: Realization of 4th Paradigm of Science in Materials Science, *SJTU/MaGIC Faculty Visit*, April 27, 2017, Evanston IL, USA.
16. (Invited) Materials Informatics and Big Data: Realization of 4th Paradigm of Science in Materials Science: Steel Fatigue Strength Predictor, *5th NU-NIMS Materials Genome Workshop*, March 28, 2017, Evanston IL, USA.
17. CHiMaD Data Mining, *CHiMaD Annual Meeting*, March 27, 2017, Evanston IL, USA.
18. (Invited) CHiMaD Data Mining: Fatigue, *SRG 2017*, March 23, 2017, Evanston IL, USA.
19. Data Science Approaches for Predicting Fatigue Strength of Steels, *TMS 2017*, February 27, 2017, San Diego CA, USA.
20. Data Science Approaches for Predicting Thermoelectric Properties, *TMS 2017*, February 27, 2017, San Diego CA, USA.
21. Parallel Implementation of Lossy Data Compression for Temporal Data Sets, *23rd Annual International Conference on High Performance Computing, Data, and Analytics (HiPC)*, December 20, 2016, Hyderabad, India.
22. Five Year Life Expectancy Calculator for Older Adults, *IEEE International Conference on Data Mining (ICDM)*, December 13, 2016, Barcelona, Spain.
23. A Formation Energy Predictor for Crystalline Materials Using Ensemble Data Mining, *IEEE International Conference on Data Mining (ICDM)*, December 13, 2016, Barcelona, Spain.
24. Predicting the Outcome of Startups: Less Failure, More Success, *IEEE ICDM Workshop on Data Market for Co-evolution of Sciences and Business (MoDAT)*, December 12, 2016, Barcelona, Spain.
25. (Invited) Materials Informatics and Big Data: Realization of the “Fourth Paradigm” of Science in Materials Science, *CHiMaD Summit on Data & Analytics for Materials Research*, November 02, 2016, Evanston IL, USA.
26. A Fatigue Strength Predictor for Steels Using Ensemble Data Mining, *25th ACM International Conference on Information and Knowledge Management (CIKM)*, October 26, 2016, Indianapolis IN, USA.
27. (Invited) High Performance Data Mining: An Essential Paradigm for Big Data Analytics and Knowledge Discovery, *3M Visit*, October 19, 2016, Evanston IL, USA.

28. Identifying HotSpots in Five Year Survival Electronic Health Records of Older Adults, *6th IEEE International Conference on Computational Advances in Bio and Medical Sciences (IC-CABS)*, October 13, 2016, Atlanta GA, USA.
29. Software tools for sequence comparison, sequence mapping, and patient-specific healthcare outcome prediction, *7th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*, October 05, 2016, Seattle WA, USA.
30. Materials Informatics and Big Data: Realization of 4th Paradigm of Science in Materials Science, *International Conference on Fatigue Damage of Structural Materials (FATD)*, September 22, 2016, Hyannis MA, USA.
31. CHiMaD Data Mining: An Update, *CHiMaD Executive Meeting*, September 12, 2016, Evanston IL, USA.
32. Materials Informatics and Big Data: Realization of 4th Paradigm of Science in Materials Science, *Theory and Applications of Computational Chemistry (TACC)*, August 30, 2016, University of Washington, Seattle WA, USA.
33. Deep Learning Based Big Data Analytics in Materials Science, *MURI Annual Meeting*, August 22, 2016, Caltech, Pasadena CA, USA.
34. (Invited) Data-driven materials science enabling large-scale property prediction and optimization, *APS/CNM Users Meeting 2016*, May 10, 2016, Argonne National Lab, Chicago IL, USA.
35. (Invited) High Performance Data Mining: An Essential Paradigm for Big Data Analytics and Knowledge Discovery, *Invited seminar*, May 05, 2016, Golden CO, USA.
36. (Invited) Big data analytics in medicine and healthcare: Analyzing electronic healthcare records, sequence data, social media, and more, *Outcomes Research Workshop*, April 27, 2016, University of Chicago, Chicago IL, USA.
37. CHiMaD Data Mining, *CHiMaD Annual Meeting*, March 23, 2016, Evanston IL, USA.
38. (Invited) CHiMaD Data Mining: Fatigue, *SRG 2016*, March 21, 2016, Evanston IL, USA.
39. (Invited) High Performance Data Mining: An Essential Paradigm for Big Data Analytics and Knowledge Discovery, *Invited seminar*, March 02, 2016, Auburn AL, USA.
40. (Invited) Towards Better Efficiency and Accuracy: Data Mining for Prediction and Optimization in Materials System Design, *TMS 2016*, February 16, 2016, Nashville TN, USA.
41. CHiMaD Data Mining: An Update, *CHiMaD Executive Meeting*, October 19, 2015, Evanston IL, USA.
42. Parallel Distributed-Memory Based Community Detection for Large Graphs, *DARPA GRAPHS / SIMPLEX Workshop: Data, Algorithms and Problems on Graphs (DAPG) 2015*, September 28, 2015, New York, NY, USA.
43. (Invited) Big Data Analytics and Discovery in Medicine and Healthcare, *NUS Surgical Faculty Visit*, September 18, 2015, Chicago Innovation Exchange, Chicago IL, USA.
44. Pruned Search: A Machine Learning Based Meta-Heuristic Approach for Constrained Continuous Optimization, *8th International Conference on Contemporary Computing (IC3) 2015*, August 20, 2015, Noida, UP, India.
45. All Your Google and Facebook Logins are Belong to Us: A Case for Single Sign-off, *8th International Conference on Contemporary Computing (IC3) 2015*, August 20, 2015, Noida, UP, India.
46. (Invited) Towards an Infrastructure for Materials Data Analytics: An Overview of Tools and Platforms for Code Development, Collaboration, and Data Analytics, *Materials Research Collaboration Environment Workshop*, August 13, 2015, Dayton OH, USA.
47. Application of Machine Learning to Materials Discovery and Development, *MURI 3-Year Review*, June 23, 2015, Arlington VA, USA.

48. Optimization of Microstructures in Magnetoelastic Alloys, *MURI 3-Year Review*, June 23, 2015, Arlington VA, USA.
49. Mining of Process-Structure-Property Linkages Using Data Science Tools, *MURI 3-Year Review*, June 23, 2015, Arlington VA, USA.
50. Towards Better Efficiency and Accuracy: Data Mining for Optimization and Prediction in Materials System Design, *AFOSR Program Review*, May 21, 2015, Arlington VA, USA.
51. CHiMaD Data Mining, *CHiMaD Annual Meeting*, May 01, 2015, Evanston IL, USA.
52. Data-driven Analytics for Materials Science: Realization of the Fourth Paradigm of Science, *DARPA SIMPLEX Pre-Kickoff Meeting*, April 14, 2015, Evanston IL, USA.
53. (Invited) CHiMaD Data Mining: Fatigue, *SRG 2015*, March 23, 2015, Evanston IL, USA.
54. (Invited) High Performance Data Mining: An Essential Paradigm for Big Data Analytics and Knowledge Discovery, *Invited seminar*, March 09, 2015, Lawrence, KS, USA.
55. CHiMaD Data Mining, *CHiMaD Executive Meeting*, December 01, 2014, Evanston IL, USA.
56. (Invited) High Performance Big Data Analytics for Data-Driven Discovery in Natural Sciences, *GBMF Symposium*, July 29, 2014, Palo Alto CA, USA.
57. (Invited) CHiMaD Data Mining, *MTL/SRG 2014*, March 24, 2014, Evanston IL, USA.
58. High Performance Big Data Clustering, *SDAV All-Hands Meeting*, February 25, 2013, Atlanta GA, USA.
59. (Invited) An Overview of Essential Concepts in Data Mining, *MURI Program Review*, January 21, 2014, Dayton OH, USA.
60. Multi-objective Optimization and Multimodal Prediction in the Design of Materials System, *MURI Program Review*, January 21, 2014, Dayton OH, USA.
61. (Invited) Data-Driven Analytics and Discovery in Medicine and Healthcare, *Research Mela*, November 23, 2013, Rush Hospital, Chicago, USA.
62. An Analysis of Variation in Hospital Billing Using Medicare Data, *KDD Workshop on Data Mining for Healthcare (DMH)*, August 11, 2013, Chicago, USA.
63. Heart Transplant Outcome Prediction using UNOS Data, *KDD Workshop on Data Mining for Healthcare (DMH)*, August 11, 2013, Chicago, USA.
64. Colon Surgery Outcome Prediction Using ACS NSQIP Data, *KDD Workshop on Data Mining for Healthcare (DMH)*, August 11, 2013, Chicago, USA.
65. High Performance Big Data Clustering, *SDAV All-Hands Meeting*, February 20, 2013, San Francisco, USA.
66. Parallel Hierarchical Clustering on Shared Memory Platforms, *International Conference on High Performance Computing (HiPC)*, December 19, 2012, Pune, India.
67. Data-driven Analytics and Applications - Realization of the Fourth Paradigm of Science, *MURI Kickoff Meeting*, October 19, 2012, Carnegie Mellon University, Pittsburgh, USA.
68. Supporting Computational Data Model Representation with High-performance I/O in Parallel netCDF, *International Conference on High Performance Computing (HiPC)*, December 19, 2011, Bangalore, India.
69. Enhancing Parallelism of Pairwise Statistical Significance Estimation for Local Sequence Alignment, *2nd HiPC Workshop on Hybrid Multi-Core Computing*, December 18, 2011, Bangalore, India.
70. Identifying HotSpots in Lung Cancer Data Using Association Rule Mining, *2nd IEEE ICDM Workshop on Biological Data Mining and its Applications in Healthcare, BioDM 2011*, December 11, 2011, Vancouver, Canada.

71. Community Dynamics and Analysis of Decadal Trends in Climate Data, *3rd IEEE ICDM Workshop on Knowledge Discovery from Climate Data, ClimKD 2011*, December 11, 2011, Vancouver, Canada.
72. Derived Distribution Points Heuristic for Fast Pairwise Statistical Significance Estimation, *ACM International Conference on Bioinformatics and Computational Biology (ACM-BCB) 2010*, August 03, 2010, Niagara, New York, USA.
73. MPIPairwiseStatSig: Parallel Pairwise Statistical Significance Estimation of Local Sequence Alignment, *HPDC ECMLS 2010*, June 22, 2010, Chicago, Illinois, USA.
74. Pairwise Statistical Significance of Local Sequence Alignment Using Substitution Matrices with Sequence-Pair-Specific Distance, *IEEE International Conference on Information Technology (ICIT 2008)*, December 20, 2008, Bhubaneswar, India.
75. Pairwise DNA Alignment with Sequence Specific Transition-Transversion Ratio Using Multiple Parameter Sets, *IEEE International Conference on Information Technology (ICIT 2008)*, December 20, 2008, Bhubaneswar, India.
76. Conservative, Non-Conservative and Average Pairwise Statistical Significance of Local Sequence Alignment, *IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2008)*, November 5, 2008, Philadelphia, PA, USA.
77. Pairwise Statistical Significance of Local Sequence Alignment Using Multiple Parameter Sets, *2nd International Workshop on Data and Text Mining in Bioinformatics (DTMBIO 2008)*, October 30, 2008, Napa Valley, CA, USA.
78. DNAlignTT: Pairwise DNA Alignment with Sequence Specific Transition-Transversion Ratio, *IEEE International Conference on EIT 2008*, May 21, 2008, Ames, IA, USA.
79. Pairwise Statistical Significance Versus Database Statistical Significance for Local Alignment of Protein Sequences, *International Symposium on Bioinformatics Research and Applications (ISBRA 2008)*, May 7, 2008, Atlanta, GA, USA.
80. Estimating Pairwise Statistical Significance of Protein Local Alignments Using a Clustering-Classification Approach Based on Amino Acid Composition, *International Symposium on Bioinformatics Research and Applications (ISBRA 2008)*, May 7, 2008, Atlanta, GA, USA.
81. Identifying Temporal Gene Networks Using Signal Processing Metrics on Time-Series Gene Expression Data, *IEEE Third International Conference on Intelligent Sensing and Information Processing*, December 14, 2005, Bangalore, India.
82. Identifying Temporal Gene Networks by Mining Gene Expression Data, *IEEE 12th International Conference on Advanced Computing and Communications (ADCOM) 2004*, December 17, 2004, Ahmedabad, India.

SOFTWARE
DEVELOPED

1. Parallel Data Clustering Algorithms: Software for parallel DBSCAN clustering in MPI and OpenMP, along with sample dataset is available at <http://cucis.ece.northwestern.edu/projects/Clustering/index.html>
2. Steel Fatigue Strength Predictor: The online steel fatigue strength predictor is available at <http://info.eecs.northwestern.edu/SteelFatigueStrengthPredictor>
3. Formation Energy Predictor: The online formation energy predictor is available at <http://info.eecs.northwestern.edu/FEpredictor>
4. ThermoEl Toolkit: The online thermoelectric toolkit for predicting Seebeck coefficient and other thermoelectric properties is available at <http://info.eecs.northwestern.edu/ThermoEl>
5. Five Year Life Expectancy Calculator: The online five year life expectancy calculator for older adults is available at <http://info.eecs.northwestern.edu/FiveYearLifeExpectancyCalculator>
6. Lung Cancer Outcome Calculator: The online lung cancer outcome calculator is available at <http://info.eecs.northwestern.edu/LungCancerOutcomeCalculator>

7. Sentiment Analysis for Social Media Data: An API for sentiment analysis service and benchmark data is available at http://cucis.ece.northwestern.edu/projects/Social/sentiment_api.html
8. Analyzing the Variation in Hospital Billing using Medicare Data: Visually depicted results in the form of US state heat maps can be found at <http://users.eecs.northwestern.edu/~ankitag/hospitalbilling/>
9. Real-Time Disease Surveillance using Social Media Data: The real-time disease surveillance tool for flu is available at <http://pulse.eecs.northwestern.edu/~kml649/flu/>
10. Poll: Identifying High-Impact Contributions of an Article: The prototype version of Poll is available at <http://info.eecs.northwestern.edu/~lpolepeddi/poll/>
11. Pairwise Statistical Significance Estimation: Sequential codes for pairwise statistical significance estimation and its variants are available at <http://www.cs.iastate.edu/~ankitag/projects.html>. Parallel codes for optimized HPC implementations (on MPI, OpenMP, Hybrid, GPU) of pairwise statistical significance estimation are available at <http://cucis.ece.northwestern.edu/projects/PSSE/>
12. AGILE: Long Read Sequence Mapping: Software and data for AGILE are available at <http://cucis.ece.northwestern.edu/projects/NGS/agile.html>
13. DNAlignTT: DNA Alignment with Sequence-Specific Transition-Transversion Ratio: Source code for DNAlignTT is available at <http://www.cs.iastate.edu/~ankitag/DNAlignTT.html>
14. FATBEP: Fuzzy Adaptive Thresholding Based Exon Predictor: An implementation of the approach as a user friendly GUI in MATLAB is available at <http://www.cs.iastate.edu/~ankitag/FATBEP.html>

RESEARCH
SUPERVISION

Post Doctoral Fellows and/or Research Associates

- Dianwei Han (2015-2017)
- Al'ona Furmanchuk (2015-2016)
- Lu Liu (2013-2014)
- Seung Woo Son (2013-2014)
- Zhengzhang Chen (2012-2014)
- Mostofa Patwary (2011-2013)
- William Hendrix (2011-2012)

Graduate Students (as Advisor/Co-advisor/Committee Co-Chair)

- Zijiang Yang, PhD (2016-present)
- Kai Yuan Hou, PhD (2016-present)
- Dipendra Jha, PhD (2016-present)
- Aagam Divyangkumar Shah, MS (2016-present)
- Arindam Paul, PhD (2015-present)
- Sunwoo Lee, PhD (2015-present)
- Qiao Kang, PhD (2015-present)
- Kaushal Jitendrakumar Patel, MS (2015-present)
- Chen Jin, PhD (2013-present)
- Amar Krishna, MS (2013-2016)
- Reda Al-Bahrani, PhD (2012-present)
- Esteban (Steve) M. Rangel, PhD (2012-present)
- Kathy Lee, PhD (2010-present)
- Ruoqian (Rosanne) Liu, PhD, 2016, Dissertation Title: "Multi-Contextual Representation and Learning with Applications in Materials Knowledge Discovery" (2012-2016)
- Yusheng (Yves) Xie, PhD, 2015, Dissertation Title: "Recommender Systems for Social Networks" (2011-2015)

- Diana Palsetia, PhD, 2015, Dissertation Title: “High Quality, Scalable Community Detection for Large Graphs” (2010-2015)
- Yu Cheng, PhD, 2015, Dissertation Title: “Mining and Understanding Professional Social Network: Challenges and Solutions” (2010-2015)

Graduate Students (Other)

- Sharifah Ummu Kulthumm, MS (2015)
- Zheng (Dobbie) Yuan, PhD (2015-2016)
- Yuhong Zhang, PhD (2009-2011)
- Daniel Honbo, PhD (2009-2014)
- Kunpeng Zhang, PhD, 2013, Dissertation Title: “Big Social Media Data Mining for Marketing Intelligence” (2010-2013)
- Sanchit Misra, PhD, 2011, Dissertation Title: “High Throughput Sequence Mapping for Next Generation DNA Sequencing” (2009-2011)

Undergraduate Students

- Pranjal Daga (2015-2015)
- Siddharth Gupta (2013-2013)
- Lalith Polepeddi (2009-2010)

Graduate Student Exam Committees

- Dipendra Jha, PhD Research Proposal (2017)
- Arindam Paul, PhD Research Proposal (2017)
- Xiaolin Li (Mechanical Engineering Major), PhD Qualifying Examination (2016)
- Reda Al-Bahrani, PhD Qualifying Examination (2015), PhD Research Proposal (2017)
- Ruoqian (Rosanne) Liu, PhD Research Proposal (2015), PhD Thesis Defense Exam (2016)
- Yusheng (Yves) Xie, PhD Research Proposal (2014), PhD Thesis Defense Exam (2015)
- Yu Cheng, PhD Research Proposal (2014), PhD Thesis Defense Exam (2015)
- Esteban (Steve) M. Rangel, PhD Qualifying Examination (2014), PhD Research Proposal (2015, 2016)
- Diana Palsetia, PhD Research Proposal (2014), PhD Thesis Defense Exam (2015)
- Chen Jin, PhD Research Proposal (2013)
- Kathy Lee, PhD Qualifying Examination (2013), PhD Research Proposal (2014), PhD Thesis Defense Exam (2017)
- Kunpeng Zhang, PhD Qualifying Examination (2012), PhD Research Proposal (2012), PhD Thesis Defense Exam (2013)
- Daniel Honbo, PhD Research Proposal (2011), PhD Thesis Defense Exam (2014)
- Sanchit Misra, PhD Research Proposal (2011), PhD Thesis Defense Exam (2011)

PROFESSIONAL SERVICE

Editorial board and organizing committees

- Associate Editor, [BMC/Springer Big Data Analytics](#)
- Guest Editor, [Integrating Materials and Manufacturing Innovation \(IMMI\)](#), Special Issue on “Using Digital Data in Materials Science and Engineering”, 2013.
- Organizing committee member, Data & Analytics for Materials Research Summit, hosted by the [National Institute of Standards and Technology](#) and its [Center of Excellence for Hierarchical Materials Design \(CHiMaD\)](#), October 31 - November 2, 2016.
- Advisory committee member, Materials Data Analytics: A Path-Finding Workshop, hosted by [ASM International's CMD Network](#) and [OSU](#) with support from [NIST](#), October 8-9, 2015.

Invited program committee member

- IEEE International Conference on Data Mining (ICDM) [2017](#), [2016](#), [2015](#), [2014](#)
- IEEE International Conference on Big Data (BigData) [2017](#), [2016](#), [2015](#), [2014](#), [2013](#)

- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2016
- IEEE International Conference on Machine Learning and Data Science (ICMLDS) 2017
- ACM International Conference on Information and Knowledge Management (CIKM) 2013, 2012
- ACM/IEEE International Conference for High Performance Computing, Networking, Storage and Analysis (SC) 2017
- IEEE International Parallel & Distributed Processing Symposium (IPDPS) 2016, 2015
- IEEE ICDM Workshop on Data Mining in Networks (DaMNet) 2017, 2016, 2015, 2014, 2013, 2012
- ACM/IEEE SC Workshop on Big Data Analytics: Challenges and Opportunities (BDAC) 2015
- IEEE IPDPS Workshop on Parallel and Distributed Processing for Computational Social Systems (ParSocial) 2018, 2017, 2016
- IEEE International Conference on Systems, Man, Cybernetics (SMC) 2015
- IEEE International Conference on Big Data Science and Engineering (BDSE) 2014, 2013

Invited grant proposal reviewer/panelist

- National Science Foundation (NSF) CISE/IIS division 2014, 2013
- Air Force Office of Scientific Research (AFOSR) 2015
- Northwestern Data Science Initiative (DSI) 2016

Invited paper reviewer

- ACM Computing Surveys
- ACM Transactions on Knowledge Discovery from Data (TKDD)
- ACS Combinatorial Science
- ARC American Institute of Aeronautics and Astronautics Journal (AIAA)
- BMC Bioinformatics
- BMC Medical Informatics and Decision Making
- BMC Research Notes
- Elsevier Calphad (Computer Coupling of Phase Diagrams and Thermochemistry)
- Elsevier Computers in Biology and Medicine (CBM)
- Elsevier Computer Methods and Programs in Biomedicine (CMPB)
- Elsevier Computer Communications (COMCOM)
- Elsevier Current Opinion in Solid State & Materials Science (COSSMS)
- Elsevier Data and Knowledge Engineering (DKE)
- Elsevier Digital Signal Processing (DSP)
- Elsevier Information Sciences
- Elsevier Journal of Computational Science (JOCS)
- Elsevier Journal of Parallel and Distributed Computing (JPDC)
- Elsevier Neural Networks (NEUNET)
- Hindawi BioMed Research International
- Hindawi Scientifica
- IEEE Transactions on Computational Biology and Bioinformatics (TCBB)
- IEEE Transactions on Emerging Topics in Computing (TETC)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- IEEE Transactions on Parallel and Distributed Systems (TPDS)
- IEEE Transactions on Very Large Scale Integration Systems (TVLSI)
- IET Healthcare Technology Letters
- Inderscience International Journal of Computer Aided Engineering and Technology (IJCAET)
- Inderscience International Journal of High Performance Computing and Networking (IJHPCN)
- MDPI Algorithms
- MDPI Entropy
- OxfordUP Bioinformatics

- Springer Data Mining and Knowledge Discovery (DMKD)
- Springer Language Resources and Evaluation (LREV)
- T&F International Journal of Parallel, Emergent and Distributed Systems (IJPEDS)
- T&F Journal of Civil Engineering and Management (JCEM)
- T&F Transport
- TMS Springer Integrating Materials and Manufacturing Innovation (IMMI)
- Wiley Concurrency and Computation: Practice and Experience (CCPE)
- World Scientific Journal of Circuits, Systems, and Computers (JCSC)
- SC 2013
- ICDM 2012
- ICDM DaMNet 2011
- IPDPS 2011
- PAKDD 2010
- Languages: C/C++, MPI/OpenMP parallel programming in C/C++, SQL, HTML, Python
- Applications/Packages: LATEX, MATLAB, R, WEKA, SPSS, RapidMiner, Mallet
- Operating Systems: Windows, Unix/Linux, Mac OS

COMPUTER
SKILLS

REFERENCES

- Dr. Alok Choudhary
Henry and Isabelle Dever Professor,
Electrical Engineering and Computer Science,
Professor, Kellogg School of Management
Director, Center for Ultra-Scale Computing and Information Security
Northwestern University,
2145 Sheridan Rd. TECH building, Evanston, IL 60208, USA.
Email: choudhar@eecs.northwestern.edu, Phone: 847-467-4129, Fax: 847-467-4144.
WWW: <http://www.eecs.northwestern.edu/~choudhar/>
- Dr. Xiaoqiu Huang
Professor,
Department of Computer Science,
Iowa State University,
226 Atanasoff Hall, Ames, IA 50011, USA.
Email: xqhuang@iastate.edu, Phone: 515-294-2432, Fax: 515-294-0258.
WWW: <http://www.cs.iastate.edu/~xqhuang/>
- Dr. Surya Kalidindi
Professor,
College of Engineering, The George W. Woodruff School of Mechanical Engineering
College of Computing, School of Computational Science and Engineering
Georgia Institute of Technology,
778 Atlantic Drive, Bunger-Henry Building, Atlanta, GA 30332, USA.
Email: surya.kalidindi@me.gatech.edu, Phone: 404-385-2886.
WWW: <http://www.me.gatech.edu/faculty/kalidindi>
- Dr. Jaishankar Raman, MD, PhD
Professor and Director,
Department of Cardiovascular-Thoracic Surgery
Rush University Medical Center,
1725 W. Harrison St., Suite 1156, Chicago, IL 60612, USA.
Email: jai_raman@rush.edu, Phone: 312-563-2762, Fax: 312-942-3666.
WWW: http://doctors.rush.edu/directory/profile.asp?setsize=10&pict_id=4550439
- Dr. Wei-keng Liao
Research Professor,
Electrical Engineering and Computer Science,
Northwestern University,
2145 Sheridan Rd. TECH building, Evanston, IL 60208, USA.

Email: wkliao@eecs.northwestern.edu, Phone: 847-491-2916, Fax: 847-491-4455.
WWW: <http://www.eecs.northwestern.edu/~wkliao/>

- Dr. Marc De Graef
Professor,
Materials Science and Engineering Department,
Carnegie Mellon University,
5000 Forbes Avenue, Pittsburgh, PA 15213, USA.
Email: degraef@cmu.edu, Phone: 412-268-8527, Fax: 412-268-7596.
WWW: <http://neon.materials.cmu.edu/degraef/>