
Jimeng Sun

Health Innovation Professor at Carle's Illinois College of Medicine,
Professor in Computer Science at
University of Illinois Urbana-Champaign, IL, USA

jimeng.sun@gmail.com <http://sunlab.org/>

SUMMARY

Dr. Sun is a Health Innovation Professor at the Computer Science Department and Carle's Illinois College of Medicine at University of Illinois Urbana Champaign. He was the **global head of AI research at IQVIA**, the health data science and clinical trial company,. He also created multiple companies in AI for biotech space including Medisyn for **synthetic data generation**, Sunstella Technology for **AI for drug discovery and development**. Before UIUC, he was an associate professor in the College of Computing at Georgia Tech (GT) and **the co-director of the Center for Health Analytics and Informatics (CHAI)** and formed a health research neighborhood leading 25 faculty members across multiple schools occupying two floors of the innovation space in GT Coda building.

His research focuses on **artificial intelligence (AI) for healthcare**, including deep learning for drug discovery, clinical trial optimization, computational phenotyping, clinical predictive modeling, treatment recommendation, and health monitoring. He was recognized as one of **the Top 100 AI Leaders in Drug Discovery and Advanced Healthcare** by Deep Knowledge Analytics. He has published over **300 papers** and has **23,934 citations, h-index 78, and i10-index 219**. He has several NSF and NIH-funded projects with leading hospitals such as **MGH, Sutter Health, Vanderbilt, Northwestern, Geisinger and Emory**. Dr. Sun collaborates with the biomedical industry including IQVIA, and multiple pharmaceutical companies on health data science.

At UIUC, he already received three NSF grants related to AI for healthcare, and two seed grants with OSF healthcare.

He completed his B.S. and M.Phil. in computer science at Hong Kong University of Science and Technology in 2002 and 2003, respectively, and his Ph.D. in computer science at Carnegie Mellon University in 2007.

EDUCATION

Carnegie Mellon University 2003 - 2007

- *Ph.D. in Computer Science* 2007

- *M.S. in Computer Science* 2006

Thesis: "Incremental Pattern Discovery on Streams, Graphs and Tensors"

Adviser: Prof. Christos Faloutsos

Hong Kong University of Science and Technology 1999 - 2003

- M.S. in Computer Science 2003

- B.S. in Computer Science (minor in Mathematics) 2002

EMPLOYMENT

[Medisyn inc](#), Champaign, IL

CEO and co-founder (Jun 2022 - PRESENT)

- Business development for synthetic data generation/simulation in healthcare
- Research development for synthetic data generation/simulation in healthcare, and predictive modeling using synthetic data

[University of Illinois Urbana-Champaign](#), Champaign, IL

Health Innovation Professor, Carle's Illinois College of Medicine (Jan 2021 - PRESENT)

Professor in Computer Science, Computer Science Department (Jan 2020 - PRESENT)

- Research: My research mainly focuses on the following areas: (1) Deep learning for drug discovery, (2) Clinical trial optimization, (3) Computational phenotyping, (4) Clinical predictive modeling, (5) Treatment recommendation, (6) Health monitoring, and (7) Health data science education
- Teach: Grad-level MS/PhD online courses in Deep learning for healthcare, big data analytics in healthcare, and MD data science bootcamp
- Collaboration: MGH-sleep medicine and seizure subtyping, Sutter Health, OSF Healthcare- AI model deployment on vaccine distribution, Drug discovery, trial optimization and clinical predictive modeling with IQVIA

[IQVIA](#), Cambridge, MA

Global head of AI research (Aprc 2021 - Dec, 2021)

- Research: Leading their research efforts on clinical trial matching, trial site selection, trial success prediction, and clinical predictive modeling. Publish on leading AI/ML venues, and improve IQVIA research reputation and leadership in AI/ML in healthcare.
- Strategy: Advise IQVIA analytic center of excellence on AI and ML strategies to ensure their AI/ML methodologies are state of the art.
- External engagement: Help IQVIA to engage with external pharma clients on various pilots and proposals.

[Sunstella Technology Corporation](#), Las Vegas, NV

President (Jan 2019 - PRESENT)

- Consult for various healthcare and biotech companies in data science, AI /ML projects
- Advise on AI/ML methodologies and strategies.
- Technology vetting on AI in health and biotech startups
- Research development of various AI solutions for drug discovery and development

[Sunstella Foundation](#), San Francisco, CA

- *President* (May 2021 - PRESENT)

- Empower new and future immigrants studying in technology areas to achieve excellence and recognition.
- Organize ML summer camp for immigrant students on ML research to improve their chance to get into the top CS PhD programs.

Georgia Institute of Technology, [Atlanta, GA](#)

Associate Professor in Computer Science 2014 - 2019

Co-Director of Center for Health Analytics and Informatics 2017 - 2019

- Research: Computational phenotyping with tensor factorization and Clinical predictive modeling with deep learning
- Teach: Big data analytics for healthcare, and Ethics in computing
- Collaboration: Created and co-directs the Center for Health Analytics and Informatics (CHAI); Co-lead health research neighborhood for 25 faculty members at CODA building; develop and coordinate discussion and planning for health neighborhood space and activities for faculties from CoC, CoE and CoS

IBM TJ Watson Research Center, [Yorktown Heights NY](#)

Research Staff Member [Yorktown Heights NY](#) 2007 - 2013

Center for Computational Health (2007 - 2013)

- Lead the research and development on the following areas: (1) predictive modeling technology for personalized disease risk assessment, (2) patient similarity technology for supporting various healthcare analytics, and (3) advanced visual analytics for exploring patient cohorts

Database and Network Analysis Team (2007 - 2009)

- Design and implement a distributed data warehouse that host IT delivery data and metrics for over thousands of accounts
- Design data analytic techniques for IT service delivery data and metrics to optimize service processes.
- Research on enterprise social networks, visualization and recommendation.

AWARDS

1. Best paper published in 2020 in "Best of Medical Informatics". IMIA Yearbook on Medical Informatics, 2021.
2. [Top 100 AI Leaders in Drug Discovery and Advanced Healthcare](#) 2019
3. SDM/IBM early career research award 2017
4. Georgia Tech IDEAS award 2015
5. Google Faculty Award, 2015
6. AMIA Distinguished Paper Semi-finalist 2015
7. Best Health Connect South Collaboration Award (UCB+GT) 2015
8. IBM Master Inventor 2013
9. IBM Research Accomplishment Award for Intelligent Care Delivery Analysis 2013
10. KDD'12 Best Poster Presentation Award

-
11. AMIA Distinguished Paper Award nominee 2010
 12. IBM Research Accomplishment Award for Service Quality Research 2009
 13. ICDM'08 Best Research Paper Award
 14. KDD'08 Dissertation Award Runner-up
 15. SDM'07 Best Research Paper

GRANTS

As Principal Investigator

1. NSF SCH 2205289, Collaborative Research: SCH: Fair Federated Representation Learning for Breast Cancer Risk Scoring with Stanford, ASU, and OSF, UIUC portion \$350,000, 09/07/22 – 09/06/26 (site PI)
2. NSF SCH 2014438, SCH:INT: Collaborative Research: Deep Sense: Interpretable Deep Learning for Zero-effort Phenotype Sensing and Its Application to Sleep Medicine, collaboration with MGH and MIT, UIUC portion \$400,000, 08/15/20 – 08/14/24 (lead PI)
3. NSF PPOSS 2028839, Collaborative Research: PPOSS: Planning: Integrated Scalable Platform for Privacy-aware Collaborative Learning and Inference, collaboration with GT, Berkeley, UIUC portion \$50,000, 10/15/20 – 10/14/21
4. NSF BIG-DATA 1838042, BigData:IA:Collaborative Research: TIMES: Tensor Factorization for Irregular Multi-source and Evolving Spatio-Temporal Data, collaboration with Emory, UIUC portion \$773,528, 10/01/18 - 09/30/22, (UIUC PI)
5. NIH R01 1R01NS107291-01, ADVANCE: Big Data and Deep Learning for the Interictal-Ictal-Injury Continuum, Massachusetts General Hospital, GT portion \$793,375, 09/11/18 - 09/10/23 (UIUC PI)
6. NIH R01, Reducing Disparities among Kidney Transplant Recipients, Emory University, as site-PI, \$753,527, 09/01/17-08/31/22 (Site PI).
7. NIH R56, Interpretable Deep Learning Model for Longitudinal Electronic Health Records and Applications to Heart Failure Prediction, Georgia Tech, \$756,093, 09/13/2017 - 08/31/2018 (PI)
8. GT/Geisinger collaboration seed grant: HeartTrack: Software to predict diagnosis, therapy and outcomes from heart images: \$100,000, 9/1/2016 – 9/1/2017 (co-PIs: Brandon Fornwalt, Geisinger, Jim Rehg, GT)
9. GT/CHOA collaboration seed grant: Phenotyping Medically Complex Patients, \$50,000, 2016
10. Industry research project: Phenotyping Epilepsy Non-responders, UCB \$312,042 12/15/2015 – 12/31/2016
11. Georgia Tech IDEAS: Innovation in Data Engineering and Science Award (PIs with Jim Rehg): \$150,000 9/1/2015 – 9/1/2016
12. Just-in-time Learning Pilot for Public Health Preparedness and Response, CDC, \$155,000, 09/19/2015-09/18/2016
13. Google faculty award: Data + Knowledge Similarity Metric Learning, \$48,000, 2015
14. Industry research project: Analytics and Connectivity Platforms Proof-of-concept, UCB \$1,076,063.73, 1/1/2014 – 12/31/2017.
15. NSF SCH: INT:#1418511 Collaborative Research: High-throughput Phenotyping on Electronic Health Records using Multi-Tensor Factorization, NSF \$2.1m total (lead PI \$640,617 + \$13,920 REU), 09/01/2014-08/31/2018.
16. Similarity-based Just-in-time Learning Pilot for Public Health Informatics, CDC, \$184,956, 09/19/2014-09/18/2015
17. Scalable Healthcare Analytics using Similarity and Temporal Analysis, ORNL, \$79,315, 12/04/2014-03/31/2016.

As Co-Principal Investigator

18. NSF XPS:FULL:DSD: #1533768 A Parallel Tensor Infrastructure (ParTI!) for Data Analysis, NSF, \$750,000, 09/01/15-08/31/19 PI: Rich Vuduc (my portion \$375,000)

TEACHING

Courses Created

1. UIUC CS598: Deep learning for healthcare (Online Master in CS), Spring 2020, Spring 2021
2. GaTech CSE6250: Big Data Analytics in Healthcare (Online Master Sci in CS), Spring 2020, Spring, Fall 2019, Spring, Fall 2018, Spring, Fall 2017, Spring, Fall 2016

Courses Taught

3. GaTech CS4001: Computing & Society, Fall 2019, Fall 2018, Fall 2017, Fall 2016, Fall 2015, Spring 2014
4. GaTech VIP team: Predictive Health, Fall 2014, Spring 2015
5. GaTech CS8001: Big data analytics in healthcare, Spring 2014

TUTORIAL

1. [KDD 21] Fenglong Ma, Muchao Ye, Junyu Luo, Cao Xiao, **Jimeng Sun**. Tutorial: Advances in Mining Heterogeneous Healthcare Data, KDD, Virtual, 2021.
2. [IJCAI 20] Marinka Zitnik, Cao Xiao, **Jimeng Sun**. Tutorial: Machine Learning for Drug Discovery, the 29th International Joint Conference on Artificial Intelligence (IJCAI 2020).
3. [ODSC 19] Cao Xiao, **Jimeng Sun**. Tutorial: Deep Learning for Healthcare, Open Data Science Conference (ODSC), San Francisco, CA, 2019.
4. [KDD 19] Cao Xiao, **Jimeng Sun**. Tutorial: Data Mining Methods for Drug Discovery and Development, KDD, Anchorage, AK, 2019.
5. [KDD 18] Edward Choi, Cao Xiao, **Jimeng Sun**. Tutorial: Deep Learning for Computational Healthcare, KDD, London, UK, 2018.

TEXTBOOK

Cao Xiao, **Jimeng Sun**. [Introduction to Deep Learning for Healthcare](#), Springer; 1st ed. 2021

BOOK CHAPTER & EDITED BOOKS

1. **Jimeng Sun**, Jie Tang: A Survey of Models and Algorithms for Social Influence Analysis. Social Network Data Analytics 2011: 177-214
2. Papadimitriou Spiros, **Jimeng Sun**, Christos Faloutsos: Dimensionality Reduction and Forecasting on Streams. Data Streams - Models and Algorithms 2007: 261-288
3. **Jimeng Sun**, Yan Liu, Jie Tang, Chid Apté: Introduction to Special Issue on Large-Scale Data Mining. TKDD 5(2): 7 (2011)
4. Gkoulalas-Divanis, Aris, Grigorios Loukides, Li Xiong, **Jimeng Sun**. Informatics methods in medical privacy. Journal of Biomedical Informatics (JBI) 50: 1-3 (2014)

OPEN-SOURCE SOFTWARE

1. [\[DeepPurpose\]](#) A deep learning based molecular modeling and prediction toolkit on drug-target interaction prediction, compound property prediction, protein-protein interaction prediction, and protein function prediction (using PyTorch). It allows easy usages (several lines of codes only) to enable biomedical scientists to leverage deep learning for drug discovery.
2. [\[Therapeutics Data Commons \(TDC\)\]](#) An open-science platform with AI/ML-ready datasets and learning tasks for therapeutics, spanning the discovery and development of safe and effective medicines. It includes 22 therapeutic tasks and 66 ML-ready benchmark datasets over 15m data points. TDC also provides an ecosystem of tools, libraries, leaderboards, and community resources, including data functions, strategies for systematic model evaluation, meaningful data splits, data processors, and molecule generation oracles. All resources are integrated and accessible via an open Python library.
3. [\[PyHealth\]](#) A python machine learning library for AI in healthcare applications aiming at integrating and streamlining the development and evaluation of predictive health modeling, thus to simplify and expedite this process for health data scientists.

PUBLICATION

Dr. Sun's research focuses on machine learning and artificial intelligence for modeling real world patient and biomedical data. He has published over 300 papers and has 23,934 citations, h-index 78, and i10-index 219. The following are the selected publications.

2022

1. [\[EMNLP 22\]](#) Zifeng Wang and **Jimeng Sun**. 2022. "PromptEHR: Conditional Electronic Healthcare Records Generation with Prompt Learning". EMNLP'22.
2. [\[EMNLP 22\]](#) Zifeng Wang, Zhenbang Wu, Dinesh Agarwal, and **Jimeng Sun**. 2022. "MedCLIP: Contrastive Learning from Unpaired Medical Images and Texts" EMNLP'22
3. [\[EMNLP 22\]](#) Zifeng Wang and **Jimeng Sun**. 2022. "Trial2Vec: Zero-Shot Clinical Trial Document Similarity Search using Self-Supervision" Findings in EMNLP'22
4. [\[NeurIPS 22\]](#) Zifeng Wang and **Jimeng Sun**. 2022. "TransTab: Learning Transferable Tabular Transformers Across Tables" NeurIPS'22
5. [\[NeurIPS 22\]](#) Zhen Lin, Shubhendu Trivedi and **Jimeng Sun**. 2022. "Conformal Prediction with Temporal Quantile Adjustments." Neural Information Processing Systems NeurIPS'22
6. [\[NeurIPS 22\]](#) Tianfan Fu*, Wenhao Gao*, Connor W. Coley, **Jimeng Sun**. 2022. "Reinforced Genetic Algorithm for Structure-based Drug Design." Neural Information Processing Systems NeurIPS'22
7. [\[NeurIPS 22\]](#) Chaoqi Yang, Cheng Qian, Navjot Singh, Cao Xiao, M Brandon Westover, Edgar Solomonik, and **Jimeng Sun**. "ATD: Augmented CP Tensor Decomposition by Self-supervision." Neural Information Processing Systems NeurIPS'22
8. [\[NeurIPS 22\]](#) Wenhao Gao*, Tianfan Fu*, **Jimeng Sun**, Connor W. Coley. 2022. "Sample Efficiency Matters: A Benchmark for Practical Molecular Optimization." Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks.
9. [\[TMLR 22\]](#) Zhen Lin, Shubhendu Trivedi and **Jimeng Sun**. 2022. "Conformal Prediction Intervals with Temporal Dependence." Transactions on Machine Learning Research 2022

10. [\[Nature Chemical Biology 22\]](#) Kexin Huang*, Tianfan Fu*, Wenhao Gao*, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, **Jimeng Sun**, Marinka Zitnik. 2022. “Artificial intelligence foundation for therapeutic science.” *Nature Chemical Biology*.
11. [\[iScience 22\]](#) Junyi Gao, Chaoqi Yang, Joerg Heintz, Scott Barrows, Elise Albers, Mary Stapel, Sara Warfield, Adam Cross, **Jimeng Sun**, and N3C consortium. 2022. “MedML: Fusing Medical Knowledge and Machine Learning Models for Early Pediatric COVID-19 Hospitalization and Severity Prediction.” *iScience* 25 (9): 104970.
12. [\[BCB 22\]](#) Zifeng Wang, and **Jimeng Sun**. 2022. “SurvTRACE: Transformers for Survival Analysis with Competing Events.” In *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 1–9. BCB '22 49. New York, NY, USA: Association for Computing Machinery.
13. [\[KDD 22\]](#) Tianfan Fu and **Jimeng Sun**. 2022. “SIPF: Sampling Method for Inverse Protein Folding.” In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 378–88. KDD '22.
14. [\[KDD 22\]](#) Tianfan Fu and **Jimeng Sun**. 2022. “Antibody Complementarity Determining Regions (CDRs) Design Using Constrained Energy Model.” In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 389–99. KDD '22.
15. [\[ICLR 22\]](#) Zifeng Wang, Shao-Lun Huang, Ercan E. Kuruoglu, **Jimeng Sun**, Xi Chen, and Yefeng Zheng. 2022. “PAC-Bayes Information Bottleneck” International Conference on Learning Representation ICLR'22
16. [\[ICLR 22\]](#) Tianfan Fu, Wenhao Gao*, Cao Xiao, Jacob Yasonik, Connor W. Coley, and **Jimeng Sun**. 2022. “Differentiable Scaffolding Tree for Molecular Optimization.” International Conference on Learning Representation ICLR'22.
17. [\[WWW 22\]](#) Junyi Gao, Cao Xiao, Lucas M Glass, **Jimeng Sun**. PopNet: Real-Time Population-Level Disease Prediction with Data Latency, The World Wide Web Conference (WWW' 22) (acceptance rate 17.7%).
18. [\[Cell Patterns 22\]](#) Tianfan Fu, Huang, Kexin, Cao Xiao, Lucas M. Glass, and **Jimeng Sun**. “HINT: Hierarchical interaction network for clinical-trial-outcome predictions.” *Cell Patterns*, 2022.
19. [\[AAAI 22\]](#) Zhen Lin, Cao Xiao, Lucas Glass, M Brandon Westover, **Jimeng Sun**. SCRIB: Set-classifier with Class-specific Risk Bounds for Blackbox Models. the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI-22) (acceptance rate 15%)

2021

20. [\[NeurIPS 21\]](#) Kexin Huang, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W Coley, Cao Xiao, **Jimeng Sun**, Marinka Zitnik. Therapeutics Data Commons: Machine Learning Datasets and Tasks for Drug Discovery and Development. The Thirty-fifth Annual Conference on Neural Information Processing Systems, 2021.
21. [\[NeurIPS 21\]](#) Zhen Lin, S. Trivedi, and **Jimeng Sun**. 2021. “Locally Valid and Discriminative Prediction Intervals for Deep Learning Models.” The Thirty-fifth Annual Conference on Neural Information Processing Systems, 2021.
22. [\[Cell Patterns 21\]](#) Huang, Kexin, Cao Xiao, Lucas M. Glass, Cathy W. Critchlow, Greg Gibson, and **Jimeng Sun**. “Machine Learning Applications for Therapeutic Tasks with Genomics Data.” *Cell Patterns*, 2021.
23. [\[Genome Medicine 21\]](#) Isgut, Monica, **Jimeng Sun**, Arshed A. Quyyumi, and Greg Gibson. “Highly Elevated Polygenic Risk Scores Are Better Predictors of Myocardial Infarction Risk Early in Life than Later.” *Genome Medicine* 13 (1): 13, 2021

24. [\[T-PAMI 2021\]](#) Spadon, Gabriel, Shenda Hong, Bruno Brandoli, Stan Matwin, Jose Fernando Rodrigues-Jr, and **Jimeng Sun**. "Pay Attention to Evolution: Time Series Forecasting with Deep Graph-Evolution Learning." IEEE Transactions on Pattern Analysis and Machine Intelligence PP (April), 2021
25. [\[KDD 21\]](#) Tianfan Fu, Cao Xiao, Cheng Qian, Lucas M Glass, **Jimeng Sun**. Probabilistic and Dynamic Molecule-Disease Interaction Modeling for Drug Discovery. The 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2021), Virtual, 2021.
26. [\[KDD 21\]](#) Chaoqi Yang, Navjot Singh, Cao Xiao, Cheng Qian, Edgar Solomonik, **Jimeng Sun**. MTC: Multiresolution Tensor Completion from Partial and Coarse Observations. The 27th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2021), Virtual, 2021.
27. [\[KDD 21\]](#) Fenglong Ma, Muchao Ye, Junyu Luo, Cao Xiao, **Jimeng Sun**. Tutorial: Advances in Mining Heterogeneous Healthcare Data, KDD, Virtual, 2021.
28. [\[IJCAI 21\]](#) Chaoqi Yang, Cao Xiao, Fenglong Ma, Lucas Glass, **Jimeng Sun**. SafeDrug: Dual Molecular Graph Encoders for Safe Drug Recommendations, The 30th International Joint Conference on Artificial Intelligence (IJCAI 2021), 2021.
29. [\[IJCAI 21\]](#) Chaoqi Yang, Cao Xiao, Lucas Glass, **Jimeng Sun**. Change Matters: Medication Change Prediction with Recurrent Residual Networks, The 30th International Joint Conference on Artificial Intelligence (IJCAI 2021), 2021.
30. [\[MLHC 21\]](#) Siddharth Biswal, Soumya Ghosh, Jon Duke, Bradley Malin, Walter Stewart, Cao Xiao, **Jimeng Sun**. EVA: Generating longitudinal electronic health records using conditional variational autoencoders. Machine Learning for Healthcare (MLHC' 21).
31. [\[ACL-Findings 21\]](#) Junyu Luo, Cao Xiao, Lucas Glass, **Jimeng Sun**, Fenglong Ma. Fusion: Towards Automated ICD Coding via Feature Compression, Findings of the Association for Computational Linguistics: ACL-IJCNLP, 2021.
32. [\[Bioinformatics 21\]](#) Yue Yu, Kexin Huang, Chao Zhang, Lucas M Glass, **Jimeng Sun**, Cao Xiao. SumGNN: multi-typed drug interaction prediction via efficient knowledge graph summarization, Bioinformatics 2021 (Impact factor: 5.61).
33. [\[MLSys 21\]](#) Yue Zhao, Xiyang Hu, Cheng Cheng, Cong Wang, Changlin Wan, Wen Wang, Jianing Yang, Haoping Bai, Zheng Li, Cao Xiao, Yunlong Wang, Zhi Qiao, **Jimeng Sun**, Leman Akoglu. SUOD: Accelerating Large-Scale Unsupervised Heterogeneous Outlier Detection. Proceedings of Machine Learning and Systems, 2021.
34. [\[ACM-BCB 21\]](#) Tianfan Fu, Cao Xiao, Kexin Huang, Lucas M Glass, **Jimeng Sun**. SPEAR: self-supervised post-training enhancer for molecule optimization, The 12th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, 2021.
35. [\[WWW 21\]](#) Chacha Chen, Junjie Liang, Fenglong Ma, Lucas Glass, **Jimeng Sun** and Cao Xiao. UNITE: Uncertainty-based Health Risk Prediction Leveraging Multi-sourced Data, The World Wide Web Conference (WWW' 21) (acceptance rate 20.6%).
36. [\[WWW 21\]](#) Nghia Hoang, Shenda Hong, Cao Xiao, Bryan Low and **Jimeng Sun**. AID: Active Distillation Machine to Leverage Pre-Trained Black-Box Models in Private Data Settings, The World Wide Web Conference (WWW' 21) (acceptance rate 20.6%).
37. [\[IEEE TKDE 21\]](#) Tianfan Fu, Cao Xiao, Lucas M. Glass and **Jimeng Sun**. MOLER: Incorporate Molecule-Level Reward to Enhance Deep Generative Model for Molecule Optimization, IEEE Transactions on Knowledge and Data Engineering, 2021 (Impact factor: 3.857).
38. [\[AAAI 21\]](#) Tianfan Fu, Cao Xiao, Xinhao Li, Lucas M. Glass, **Jimeng Sun**. MIMOSA: Multi-constraint Molecule Sampling for Molecule Optimization, Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), 2021 (acceptance rate 21% (1692/9034))
39. [\[AAAI 21\]](#) Nikos Kargas, Cheng Qian, Nicholas Sidiropoulos, Cao Xiao, Lucas M. Glass, **Jimeng Sun**. STELAR: Spatio-temporal Tensor Factorization with Latent Epidemiological Regularization,

Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21), 2021 (acceptance rate 21% (1692/9034))

40. [JAMIA 21] Junyi Gao, Rakshith Sharma, Cheng Qian, Lucas M. Glass, Jeffrey Spaeder, Justin Romberg, **Jimeng Sun**, and Cao Xiao. STAN: Spatio-Temporal Attention Network for Pandemic Prediction Using Real-World Evidence, Journal of the American Medical Informatics Association 2021 (Impact factor: 4.29).

2020

41. [Nature Scientific Reports 20] Kexin Huang, Cao Xiao, Lucas M. Glass, Marinka Zitnik, **Jimeng Sun**. SkipGNN: Predicting Molecular Interactions with Skip-Graph Networks, Scientific Reports, 2020 (Impact factor: 4.259).
42. [Bioinformatics 20] Kexin Huang, Tianfan Fu, Lucas M. Glass, Marinka Zitnik, Cao Xiao, and **Jimeng Sun**. DeepPurpose: A Deep Learning Library for Drug-Target Interaction Prediction, Bioinformatics 2020 (Impact factor: 5.61).
43. [JAMIA 20] Zhang, Ziqi, Chao Yan, Thomas A. Lasko, **Jimeng Sun**, and Bradley A. Malin. "SynTEG: A Framework for Temporal Structured Electronic Health Data Simulation." Journal of the American Medical Informatics Association: JAMIA, 2020. (**Best Medical Informatics paper by IMIA**)
44. [BIBM 20] Tianfan Fu, Cao Xiao, Lucas Glass and **Jimeng Sun**. α -MOP: Molecule Optimization with α -divergence, International Conference on Bioinformatics and Biomedicine (BIBM) 2020 (acceptance rate 19.4%).
45. [JAMIA 20] Zhi Qiao, Austin Bae, Cao Xiao, Lucas Glass and **Jimeng Sun**. FLANNEL: Focal Loss Based Neural Network Ensemble for COVID-19 Detection, Journal of the American Medical Informatics Association 2020 (Impact factor: 4.29).
46. [Bioinformatics 20] Kexin Huang, Cao Xiao, Lucas Glass and **Jimeng Sun**. MolTrans: Molecular Interaction Transformer for Drug Target Interaction Prediction, Bioinformatics 2020 (Impact factor: 5.61).
47. [KDD 20] Junyi Gao, Cao Xiao, Lucas M. Glass, **Jimeng Sun**. COMPOSE: Cross-Modal Pseudo-Siamese Network for Patient Trial Matching, The 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2020), 2020.
48. [KDD 20] Kejing Yin, Ardavan Afshar, Joyce C. Ho, William K. Cheung, Chao Zhang and **Jimeng Sun**. "LogPar: Logistic PARAFAC2 Factorization for Temporal Binary Data with Missing Values." In Proceedings of KDD, 2020, 1625-1635
49. [KDD 20] Shenda Hong, Yanbo Xu, Alind Khare, Satria Priambada, Kevin O. Maher, Alaa Aljiffry, **Jimeng Sun** and Alexey Tumanov. "HOLMES: Health OnLine Model Ensemble Serving for Deep Learning Models in Intensive Care Units." In Proceedings of KDD, 2020, 1614-1624
50. [KDD 20] Yue Yu, Yinghao Li, Jiaming Shen, Hao Feng, **Jimeng Sun** and Chao Zhang "STEAM: Self-Supervised Taxonomy Expansion with Mini-Paths." In Proceedings of KDD, 2020, 1026-1035
51. [IJCAI 20] Marinka Zitnik, Cao Xiao, **Jimeng Sun**. Tutorial: Machine Learning for Drug Discovery, the 29th International Joint Conference on Artificial Intelligence (IJCAI 2020).
52. [Comput. Biol. Med 20] Shenda Hong, Yuxi Zhou, Junyuan Shang, Cao Xiao, **Jimeng Sun**. Opportunities and Challenges of Deep Learning Methods for Electrocardiogram Data: A Systematic Review, Computers in Biology and Medicine, 2020 (Impact factor: 2.286).
53. [IEEE TKDE 20] Cao Xiao, Trong Nghia Hoang, Shenda Hong, Tengfei Ma and **Jimeng Sun**. CHEER: Rich Model Helps Poor Model via Knowledge Infusion, IEEE Transactions on Knowledge and Data Engineering, 2020 (Impact factor: 3.857).

-
54. [JBI 20] Perros, Ioakeim, Xiaowei Yan, J. B. Jones, **Jimeng Sun**, and Walter F. Stewart. "Using the PARAFAC2 Tensor Factorization on EHR Audit Data to Understand PCP Desktop Work." *Journal of Biomedical Informatics* 101, 2020.
 55. [JAMIA 20] Junyi Gao, Cao Xiao, Lucas M. Glass, **Jimeng Sun**, Dr. Agent: Clinical Predictive Model via Mimicked Second Opinions, *Journal of the American Medical Informatics Association (JAMIA)*, 2020 (Impact factor: 4.29).
 56. [WWW 20] Junyi Gao, Cao Xiao, Yasha Wang, Wen Tang, Lucas M. Glass, **Jimeng Sun**. StageNet: Stage-Aware Neural Networks for Health Risk Prediction, *The World Wide Web Conference (WWW' 20)* (acceptance rate 19%).
 57. [WWW 20] Siddharth Biswal, Cao Xiao, Lucas M. Glass, Brandon Westover, and **Jimeng Sun**. CLARA: Clinical Report Auto-completion, *The World Wide Web Conference (WWW' 20)* (acceptance rate 19%).
 58. [WWW 20] Rahul Daggal, Scott Freitas, Cao Xiao, Duen Horng Chau, and **Jimeng Sun**. REST: Robust and Efficient Neural Networks for Sleep Staging in the Wild, *The World Wide Web Conference (WWW' 20)* (acceptance rate 19%).
 59. [WWW 20] Xingyao Zhang, Cao Xiao, Lucas M. Glass, and **Jimeng Sun**. DeepEnroll: Patient-Trial Matching with Deep Embedding and Entailment Prediction, *The World Wide Web Conference (WWW' 20)*(acceptance rate 19%).
 60. [AAAI 20] Kexin Huang, Cao Xiao, Nghia Hoang, Lucas Glass and **Jimeng Sun**. CASTER: Predicting Drug Interaction with Chemical Substructure Representation, *AAAI 2020* (acceptance rate 20.6%).
 61. [AAAI 20] Tianfan Fu, Cao Xiao, and **Jimeng Sun**. CORE: Automatic Molecule Optimization using Copy and Refine Strategy, *AAAI 2020* (acceptance rate 20.6%).
 62. [AAAI 20] Siddharth Biswal, Cao Xiao, Lucas Glass, Elizabeth Milkovits, and **Jimeng Sun**. Doctor2Vec: Dynamic Doctor Representation Learning for Clinical Trial Recruitment, *AAAI 2020* (acceptance rate 20.6%).
 63. [AAAI 20] Limeng Cui, Siddharth Biswal, Lucas Glass, Greg Lever, **Jimeng Sun**, and Cao Xiao. CONAN: Complementary Pattern Augmentation for Rare Disease Detection, *AAAI 2020* (acceptance rate 20.6%).

2019

64. [ACM-BCB 19] Tianfan Fu, Tian Gao, Cao Xiao, Tengfei Ma and **Jimeng Sun**. PEARL: Prototype Learning via Rule Lists, *The 10th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics* (acceptance rate 26.1%).
65. [CIRCULATION 19] Chen Robert, Walter F Stewart, **Jimeng Sun**, Kenney Ng, and Xiaowei Yan. "Recurrent Neural Networks for Early Detection of Heart Failure From Longitudinal Electronic Health Record Data." *Circulation. Cardiovascular Quality and Outcomes* 12 (10): e005114. 2019
66. [MLHC 19] Siddharth Biswal, Cao Xiao, Brandon Westover, and **Jimeng Sun**. EEGtoText: Learning to Write Medical Reports from EEG Recordings, *Machine Learning for Healthcare 2019* (acceptance rate 30.9%).
67. [MLHC 19] Irfan Al-Hussaini , Cao Xiao, Brandon Westover, and **Jimeng Sun**. SLEEPER: interpretable Sleep staging via Prototypes from Expert Rules, *Machine Learning for Healthcare 2019* (acceptance rate 30.9%).
68. [IJCAI 19] Tianfan Fu, Trong Nghia Hoang, Cao Xiao, and **Jimeng Sun**. DDL: Deep Dictionary Learning for Predictive Phenotyping, *The 28th International Joint Conference on Artificial Intelligence (IJCAI 2019)*, 2019.

69. [IJCAI 19] Junyuan Shang, Tengfei Ma, Cao Xiao, and **Jimeng Sun**. Pre-training of Graph Augmented Transformers for Medication Recommendation, The 28th International Joint Conference on Artificial Intelligence (IJCAI 2019), 2019.
70. [IJCAI 19] Shenda Hong, Cao Xiao, Tengfei Ma, Hongyan Li and **Jimeng Sun**. MINA: Multilevel Knowledge-Guided Attention for Modeling Electrocardiography Signals, The 28th International Joint Conference on Artificial Intelligence (IJCAI 2019), 2019.
71. [IJCAI 19] Shengda Hong, Cao Xiao, Tengfei Ma, Hongyan Li and **Jimeng Sun**. RDPD: Rich Data Helps Poor Data via Imitation, The 28th International Joint Conference on Artificial Intelligence (IJCAI 2019), 2019.
72. [KDD 19] Cao Xiao, **Jimeng Sun**. Tutorial: Data Mining Methods for Drug Discovery and Development, KDD, Anchorage, AK, 2019.
73. [WWW 19] Sungtae An, Cao Xiao, Walter Stewart, and **Jimeng Sun**. LAVA: Longitudinal Adversarial Attack on Electronic Health Records Data, The 2019 Web Conference (WWW 2019) (acceptance rate 20%), 2019.
74. [AAAI 19] Junyuan Shang, Cao Xiao, Tengfei Ma, Hongyan Li and **Jimeng Sun**. GAMENet: Graph Augmented MEmory Networks for Recommending Medication Combination, The 33rd AAAI Conference on Artificial Intelligence (AAAI 2019) (acceptance rate 16.2%), 2019.

2018

75. [NeurIPS 18] Edward Choi, Cao Xiao, Walter Stewart, and **Jimeng Sun**. MiME: Multilevel Medical Embedding from Electronic Health Records for Predictive Healthcare, \textit{The Thirty-second Annual Conference on Neural Information Processing Systems}, 2018.
76. [KDD 18] Edward Choi, Cao Xiao, **Jimeng Sun**. Tutorial: Deep Learning for Computational Healthcare, KDD, London, UK, 2018.
77. [JAMIA 18] Cao Xiao, Edward Choi, **Jimeng Sun**, Opportunities and Challenges in Developing Deep Learning Models Using Electronic Health Records Data: a systematic review, Journal of the American Medical Informatics Association (JAMIA), Oct, 2018 (Impact factor: 4.29) (Editor's Choice).
78. [JAMIA 18] Biswal, Siddharth, Haoqi Sun, Balaji Goparaju, M. Brandon Westover, **Jimeng Sun**, and Matt T. Bianchi. "Expert-Level Sleep Scoring with Deep Neural Networks." Journal of the American Medical Informatics Association: JAMIA, November. <https://doi.org/10.1093/jamia/ocy131>. 2018
79. [JMIR 18] Huang, Yingxiang, Junghye Lee, Shuang Wang, **Jimeng Sun**, Hongfang Liu, and Xiaoqian Jiang. "Privacy-Preserving Predictive Modeling: Harmonization of Contextual Embeddings From Different Sources." JMIR Medical Informatics 6 (2): e33, 2018.
80. [SC 18] Li, Jiajia, **Jimeng Sun**, and Richard Vuduc, HiCOO: Hierarchical Storage of Sparse Tensors, ACM/IEEE International Conference for High-Performance Computing, Networking, Storage, and Analysis (SC). 2018 (Best Student Paper)
81. [CIKM 18] Afshar, Ardavan, Ioakeim Perros, Evangelos E. Papalexakis, Elizabeth Searles, Joyce Ho, and **Jimeng Sun**. "COPA: Constrained PARAFAC2 for Sparse & Large Datasets." CIKM 2018
82. [KDD 18] Perros, Ioakeim, Evangelos E. Papalexakis, Haesun Park, Richard W. Vuduc, Xiaowei Yan, Christopher deFilippi, Walter F. Stewart and **Jimeng Sun** "SUSTain: Scalable Unsupervised Scoring for Tensors and its Application to Phenotyping." In Proceedings of KDD, 2018, 2080-2089
83. [KDD 18] Xu, Yanbo, Siddharth Biswal, Shriprasad R. Deshpande, Kevin O. Maher and **Jimeng Sun**. "RAIM: Recurrent Attentive and Intensive Model of Multimodal Patient Monitoring Data." In Proceedings of KDD, 2018, 2565-2573
84. [NAACL-HLT 18] James Mullenbach, Sarah Wiegrefe, Jon Duke, **Jimeng Sun** and Jacob Eisenstein. "Explainable Prediction of Medical Codes from Clinical Text." In Proceedings of NAACL-HLT, 2018,

2017

85. [KDD 17] Inci Baytas, **Cao Xiao**, Xi Zhang, Fei Wang, Anil Jain, and Jiayu Zhou. Patient-subtyping via Time-aware LSTM Networks, The 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2017) (Oral, acceptance rate 8.56%)
86. [AMIA 17] Xu, Yanbo, Mohammad Taha Bahadori, Elizabeth Searles, Michael Thompson, Javier Tejedor-Sojo, and **Jimeng Sun**. "Predicting Changes in Pediatric Medical Complexity using Large Longitudinal Health Records" AMIA 2017.
87. [AMIA 17] Deng, Yu, Al'ona Furmanchuk, Robert Chen, Faraz S. Ahmad, **Jimeng Sun**, and Abel Kho. "Use of Clinical Phenotypes and Non-negative Tensor Factorization for Heart Failure Prediction." AMIA 2017.
88. [KDD 17] Perros, Ioakeim, Evangelos E. Papalexakis, Fei Wang, Richard Vuduc, Elizabeth Searles, Michael Thompson, and **Jimeng Sun**. "SPARTan: Scalable PARAFAC2 for Large & Sparse Data." KDD, 2017.
89. [KDD 17] Zhang, Yutao, Robert Chen, Jie Tang, Walter F. Stewart, and **Jimeng Sun**. "LEAP: Learning to Prescribe Effective and Safe Treatment Combinations for Multimorbidity." KDD, 2017.
90. [KDD 17] Choi, Edward, Mohammad Taha Bahadori, Le Song, Walter F. Stewart, and **Jimeng Sun**. "GRAM: Graph-Based Attention Model for Healthcare Representation Learning." KDD, 2017.
91. [KDD 17] Kim, Yejin, **Jimeng Sun**, Hwanjo Yu, and Xiaoqian Jiang. "Federated Tensor Factorization for Computational Phenotyping." KDD, 2017.
92. [AAAI 17] Zhang, Jing, Jie Tang, Yuanyi Zhong, Yuchen Mo, Juanzi Li, Guojie Song, Wendy Hall, and **Jimeng Sun**. "StructInf: Mining Structural Influence from Social Streams." In Thirty-First AAAI Conference on Artificial Intelligence (AAAI) 2017.
93. [WWW 17] Zhang, Han, Maosong Sun, Xiaochen Wang, Zhengyang Song, Jie Tang, and **Jimeng Sun**. 2017. "Smart Jump: Automated Navigation Suggestion for Videos in MOOCs." In Proceedings of the 26th International Conference on World Wide Web (Companion Volume), 331–39.

2016 and before

94. [JAMIA 16] Choi, Edward, Andy Schuetz, Walter F Stewart, and **Jimeng Sun** "Using recurrent neural network models for early detection of heart failure onset" Journal of the American Medical Informatics Association (JAMIA) 2016
95. [AIM 16] RL Richesson, **Jimeng Sun**, J Pathak, AN Kho, and JC Denny, A survey of clinical phenotyping in selected national networks: demonstrating the need for high-throughput, portable, and computational methods, Artificial Intelligence in Medicine 71, 57-61* (impact factor 2.142)
96. [JBHI 16] Chen, Robert, **Jimeng Sun**, Robert S. Dittus, Daniel Fabbri, Jacqueline Kirby, Cheryl L. Laffer, Candace D. McNaughton, and Bradley Malin. 2016. "Patient Stratification Using Electronic Health Records from a Chronic Disease Management Program." IEEE Journal of Biomedical and Health Informatics, January 2016. doi:10.1109/JBHI.2016.2514264. (impact factor 2.093)
97. [JBI 16] K Malhotra, SB Navathe, DH Chau, C Hadjipanayis, and **Jimeng Sun**, "Constraint based temporal event sequence mining for Glioblastoma survival prediction", Journal of biomedical informatics (JBI) 61, 267-275, 2016
98. [NeurIPS 16] Choi, Edward, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, and **Jimeng Sun**. "RETAIN: Interpretable Predictive Model in Healthcare Using Reverse Time Attention Mechanism." NeurIPS'16
99. [KDD 16] Choi, Edward, Mohammad Bahadori, Elizabeth Searles, Catherine Coffey, Michael Thompson, James Bost, Javier Tejedor-Sojo, and **Jimeng Sun**. "Multi-layer Representation Learning for Medical Concepts" KDD 2016.

100. [KDD 16] Zhang, Yuyu, Mohammad Bahadori, Hang Su, and **Jimeng Sun**, “FLASH: Fast Bayesian Optimization for Data Analytic Pipelines”, KDD 2016.
101. [WSDM 16] Wu, Zhaoming, Charu C Aggarwal, **Jimeng Sun**, “The Troll-Trust Model for Ranking in Signed Networks”, Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM), 447-456, 2016
102. [ICWSM 16] Yin, Zhijun, You Chen, Daniel Fabbri, **Jimeng Sun**, and Bradley Malin. 2016. “#PrayForDad: Learning the Semantics Behind Why Social Media Users Disclose Health Information.” In Tenth International AAAI Conference on Web and Social Media. (ICWSM) 456-465
103. [MLHC 16] Edward Choi, Mohammad Taha Bahadori, Andy Schuetz, Walter F. Stewart, and **Jimeng Sun**, “Doctor AI: Predicting Clinical Events via Recurrent Neural Networks”, Machine learning for Healthcare (MLHC) 2016
104. [ICDM 15] Ioakeim Perros, Robert Chen, Richard W. Vuduc, **Jimeng Sun**: “Sparse Hierarchical Tucker Factorization and Its Application to Healthcare”. ICDM 2015: 943-948
105. [AMIA 15] Khalilia, Mohammed, Myung Choi, Amelia Henderson, Sneha Iyengar, Mark L. Braunstein and **Jimeng Sun**. “Clinical Predictive Modeling Development and Deployment through FHIR Web Services”, AMIA 2015
106. [JAMIA 15] Han Dong, Shuang Wang, Chao Jiang, Xiaoqian Jiang, Hyeon-Eui Kim, **Jimeng Sun**, and Lucila Ohno-Machado. “Trends in Biomedical Informatics: Automated Topic Analysis of JAMIA Articles.” Journal of the American Medical Informatics Association (JAMIA) 22, no. 6 1153–63. doi:10.1093/jamia/ocv157, 2015
107. [JDIQ 2015] Basole, Rahul C., Mark L. Braunstein, and **Jimeng Sun**. “Data and Analytics Challenges for a Learning Healthcare System.” J. Data and Information Quality 6, no. 2–3 (2015)
108. [JBI 15] Chen, You, Joydeep Ghosh, Cosmin Adrian Bejan, Carl A. Gunter, Siddharth Gupta, Abel N. Kho, David M. Liebovitz, **Jimeng Sun**, Joshua C. Denny, and Bradley Malin. “Building Bridges across Electronic Health Record Systems through Inferred Phenotypic Topics.” Journal of Biomedical Informatics (JBI) 55 (2015): 82–93. doi:10.1016/j.jbi.2015.03.011.
109. [JBHI 15] Fei Wang and **Jimeng Sun**. “PSF: A Unified Patient Similarity Evaluation Framework Through Metric Learning With Weak Supervision.” IEEE J. Biomedical and Health Informatics 19, no. 3 (2015): 1053–60. doi:10.1109/JBHI.2015.2425365. (impact factor 2.093)
110. [DMKD 15] Fei Wang and **Jimeng Sun**. “Survey on Distance Metric Learning and Dimensionality Reduction in Data Mining.” Data Min. Knowl. Discov. 29, no. 2 (2015): 534–64.
111. [AMIA 15] Chen, Robert, Hang Su, Yi Zhen, Mohammed Khalilia, Daniel Hirsch, Michael Thompson, Tod Davis, Yue Peng, Sizhe Lin, Javier Tejedor-Sojo, Elizabeth Searles and **Jimeng Sun**. “Cloud-based Predictive Modeling System and its Application to Asthma Readmission Prediction”, AMIA 2015
112. [SC 15] Li, Jiajia, Casey Battaglino, Ioakeim Perros, **Jimeng Sun**, and Richard W. Vuduc. “An Input-Adaptive and in-Place Approach to Dense Tensor-Times-Matrix Multiply.” In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, SC 2015, Austin, TX, USA, November 15-20, 2015
113. [AMIA 15] Ng, Kenney, **Jimeng Sun**, Jianying Hu, and Fei Wang. “Personalized Predictive Modeling and Risk Factor Identification Using Patient Similarity.” AMIA Joint Summits on Translational Science Proceedings AMIA Summit on Translational Science 2015 (2015): 132–36.
114. [KDD 15] Wang, Yichen, Robert Chen, Joydeep Ghosh, Joshua C. Denny, Abel Kho, You Chen, Bradley A. Malin, and **Jimeng Sun**. “Rubik: Knowledge Guided Tensor Factorization and Completion for Health Data Analytics.” KDD ’15.
115. [KDD 14] Ho, Joyce C., Joydeep Ghosh, and **Jimeng Sun**. “Marble: High-Throughput Phenotyping from Electronic Health Records via Sparse Nonnegative Tensor Factorization.” KDD 2014.

-
116. [JBI 14] Ho, Joyce C., Joydeep Ghosh, Steven R. Steinhubl, Walter F. Stewart, Joshua C. Denny, Bradley A. Malin, and **Jimeng Sun**. "Limestone: High-Throughput Candidate Phenotype Generation via Tensor Factorization." *Journal of Biomedical Informatics (JBI)* 52 (2014): 199–211.
 117. [JCF 14] Vijayakrishnan, Rajakrishnan, Steven R. Steinhubl, Kenney Ng, **Jimeng Sun**, Roy J. Byrd, Zahra Daar, Brent A. Williams, Christopher Defilippi, Shahram Ebadollahi, and Walter F. Stewart. "Prevalence of Heart Failure Signs and Symptoms in a Large Primary Care Population Identified Through the Use of Text and Data Mining of the Electronic Health Record." *Journal of Cardiac Failure*, April 4, 2014. doi:10.1016/j.cardfail.2014.03.008. (impact factor 3.259)
 118. [JBI 14] Gkoulalas-Divanis, Aris, Grigorios Loukides, and **Jimeng Sun**. "Publishing Data from Electronic Health Records While Preserving Privacy: A Survey of Algorithms." *Journal of Biomedical Informatics (JBI)*, June 14, 2014. doi:10.1016/j.jbi.2014.06.002.*
 119. [IBM JRD 14] Gkoulalas-Divanis, Aris, Grigorios Loukides, and **Jimeng Sun**. "Toward Smarter Healthcare: Anonymizing Medical Data to Support Research Studies." *IBM Journal of Research and Development* 58, no. 1 (2014). doi:10.1147/JRD.2013.2288173.
 120. [JBI 14] Ng, Kenney, Amol Ghoting, Steven R. Steinhubl, Walter F. Stewart, Bradley Malin, and **Jimeng Sun**. "PARAMO: A PARAllel Predictive MOdeling Platform for Healthcare Analytic Research Using Electronic Health Records." *Journal of Biomedical Informatics (JBI)* 48 (April 2014): 160–70. doi:10.1016/j.jbi.2013.12.012.
 121. [BIH 14] Ho, Joyce C., Joydeep Ghosh, and Jimeng Sun. "Extracting Phenotypes from Patient Claim Records Using Nonnegative Tensor Factorization." In *Brain Informatics and Health - International Conference, BIH 2014, Warsaw, Poland, August 11-14, 2014, Proceedings*,
 122. [JAMIA 14] **Jimeng Sun**, Candace D. McNaughton, Ping Zhang, Adam Perer, Aris Gkoulalas-Divanis, Joshua C. Denny, Jacqueline Kirby, Thomas Lasko, Alexander Saip, and Bradley A. Malin. "Predicting Changes in Hypertension Control Using Electronic Health Records from a Chronic Disease Management Program." *Journal of the American Medical Informatics Association (JAMIA)* 21, no. 2 (April 2014): 337–44. doi:10.1136/amiajnl-2013-002033.
 123. [JCF 14] Vijayakrishnan, Rajakrishnan, Steven R. Steinhubl, Kenney Ng, **Jimeng Sun**, Roy J. Byrd, Zahra Daar, Brent A. Williams, Christopher Defilippi, Shahram Ebadollahi, and Walter F. Stewart. "Prevalence of Heart Failure Signs and Symptoms in a Large Primary Care Population Identified Through the Use of Text and Data Mining of the Electronic Health Record." *Journal of Cardiac Failure*, April 4, 2014. doi:10.1016/j.cardfail.2014.03.008. (impact factor 3.259)
 124. [ICDM 13] Jiayu Zhou, Yashu Liu, **Jimeng Sun**, Hu, Jianying, and Jieping Ye. "Patient Risk Prediction Model via Top-K Stability Selection." *ICDM 2013*
 125. [KDD 13] Tang, Jie, Sen Wu, and **Jimeng Sun**. "Confluence: Conformity Influence in Large Social Networks." *KDD 2013*.
 126. [WSDM 13] Wu, Sen, **Jimeng Sun**, and Jie Tang. "Patent Partner Recommendation in Enterprise Social Networks." *Sixth ACM International Conference on Web Search and Data Mining, 2013*.
 127. [KDD 13] Zhou, Jiayu, Zhaosong Lu, **Jimeng Sun**, Lei Yuan, Fei Wang, and Jieping Ye. "FeaFiner: Biomarker Identification from Medical Data through Feature Generalization and Selection." *KDD 2013*.
 128. [JMI 13] Byrd, Roy J., Steven R. Steinhubl, **Jimeng Sun**, Shahram Ebadollahi, and Walter F. Stewart. "Automatic Identification of Heart Failure Diagnostic Criteria, Using Text Analysis of Clinical Notes from Electronic Health Records." *International Journal of Medical Informatics*, January 10, 2013. doi:10.1016/j.ijmedinf.2012.12.005. (impact factor 2.363)
 129. [AMIA 12] Gotz, David, Harry Stavropoulos, **Jimeng Sun**, and Fei Wang. "ICDA: A Platform for Intelligent Care Delivery Analytics." *AMIA* (2012): 264–73.

130. [\[AMIA 12\]](#) Hu, Jianying, Fei Wang, **Jimeng Sun**, Robert Sorrentino, and Shahram Ebadollahi. "A Healthcare Utilization Analysis Framework for Hot Spotting and Contextual Anomaly Detection." AMIA 2012.
131. [\[T-PAMI 12\]](#) Fei Wang, Noah Lee, Jianying Hu, **Jimeng Sun**, Shahram Ebadollahi, and Andrew F. Laine. "A Framework for Mining Signatures from Event Sequences and Its Applications in Healthcare Data." IEEE Transactions on Pattern Analysis and Machine Intelligence 35, no. 2 (February 2013): 272–85. doi:10.1109/TPAMI.2012.111. (impact factor 6.077)
132. [\[IBM JRD 12\]](#) Gotz, David, **Jimeng Sun**, and Nan Cao. "Multifaceted Visual Analytics for Healthcare Applications." IBM Journal of Research and Development 56, no. 5 (2012): 6.
133. [\[VLDB 12\]](#) Kang, U., Hanghang Tong, **Jimeng Sun**, Ching-Yung Lin, and Christos Faloutsos. "Gbase: An Efficient Analysis Platform for Large Graphs." VLDB J. 21, no. 5 (2012): 637–50. doi:10.1007/s00778-012-0283-9. (impact factor 1.744)
134. [\[IBM JRD 12\]](#) Markatou, Marianthi, Prabani Kuruppumullage Don, Jianying Hu, Fei Wang, **Jimeng Sun**, Robert Sorrentino, and Shahram Ebadollahi. "Case-Based Reasoning in Comparative Effectiveness Research." IBM Journal of Research and Development 56, no. 5 (2012)
135. [\[JAMIA 12\]](#) Sondhi, Parikshit, **Jimeng Sun**, ChengXiang Zhai, Robert Sorrentino, and Martin S. Kohn. "Leveraging Medical Thesauri and Physician Feedback for Improving Medical Literature Retrieval for Case Queries." Journal of the American Medical Informatics Association (JAMIA) 19, no. 5 (October 2012): 851–58. doi:10.1136/amiajnl-2011-000293.
136. [\[IBM JRD 12\]](#) Daby Sow., **Jimeng Sun**, Alain Biem, Jianying Hu, Marion Blount, and Shahram Ebadollahi. "Real-Time Analysis for Short-Term Prognosis in Intensive Care." IBM Journal of Research and Development 56, no. 5 (2012): 3. doi:10.1147/JRD.2012.2197952.
137. [\[SIGKDD Explorations 12\]](#) **Jimeng Sun**, Fei Wang, Jianying Hu, and Shahram Ebadollahi. "Supervised Patient Similarity Measure of Heterogeneous Patient Records." SIGKDD Explorations 14, no. 1 (2012): 16–24.
138. [\[TAC 12\]](#) Jie Tang, Yuan Zhang, **Jimeng Sun**, Jinghai Rao, Wenjing Yu, Yiran Chen, and Alvis Cheuk M. Fong. "Quantitative Study of Individual Emotional States in Social Networks." Transactions on Affective Computing 3, no. 2 (2012): 132–44. (impact factor 1.873)
139. [\[SADM 12\]](#) Fei Wang, **Jimeng Sun**, and Shahram Ebadollahi. "Composite Distance Metric Integration by Leveraging Multiple Experts' Inputs and Its Application in Patient Similarity Assessment." Statistical Analysis and Data Mining 5, no. 1 (2012): 54–69. doi:10.1002/sam.11135.
140. [\[SDM 12\]](#) Kang, U., Hanghang Tong, and **Jimeng Sun**. "Fast Random Walk Graph Kernel." SDM 2012.
141. [\[SDM 12\]](#) Luo, Dijun, Fei Wang, **Jimeng Sun**, Marianthi Markatou, Jianying Hu, and Shahram Ebadollahi. "SOR: Scalable Orthogonal Regression for Low-Redundancy Feature Selection and Its Healthcare Applications." SDM 2012.
142. [\[KDD 12\]](#) Miao, Gengxin, Ziyu Guan, Louise E. Moser, Xifeng Yan, Shu Tao, Nikos Anerousis, and **Jimeng Sun**. "Latent Association Analysis of Document Pairs." KDD 2012.
143. [\[AMIA 12\]](#) Perer, Adam, and **Jimeng Sun**. "MatrixFlow: Temporal Network Visual Analytics to Track Symptom Evolution during Disease Progression." AMIA 2012 (2012): 716–25.
144. [\[KDD 12\]](#) Sondhi, Parikshit, **Jimeng Sun**, Hanghang Tong, and ChengXiang Zhai. "SympGraph: A Framework for Mining Clinical Notes through Symptom Relation Graphs." KDD 2012.
145. [\[AMIA 12\]](#) **Jimeng Sun**, Jianying Hu, Dijun Luo, Marianthi Markatou, Fei Wang, Shahram Ebadollahi, Steven E. Steinhubl, Zahra Daar, and Walter F. Stewart. "Combining Knowledge and Data Driven Insights for Identifying Risk Factors Using Electronic Health Records." AMIA 2012 (2012): 901–10.
146. [\[KDD 12\]](#) Tang, Jie, Sen Wu, **Jimeng Sun**, and Hang Su. "Cross-Domain Collaboration Recommendation." KDD 2012.

-
147. [KDD 12] Wang, Fei, Noah Lee, Jianying Hu, **Jimeng Sun**, and Shahram Ebadollahi. "Towards Heterogeneous Temporal Clinical Event Pattern Discovery: A Convolutional Approach." KDD 2012.
 148. [ICPR 12] Wang, Fei, Jianying Hu, and **Jimeng Sun**. "Medical Prognosis Based on Patient Similarity and Expert Feedback." In Proceedings of the 21st International Conference on Pattern Recognition, ICPR 2012.
 149. [HISB 11] Lee, Noah, Andrew F. Laine, Jianying Hu, Fei Wang, **Jimeng Sun**, and Shahram Ebadollahi. "Mining Electronic Medical Records to Explore the Linkage between Healthcare Resource Utilization and Disease Severity in Diabetic Patients." In 2011 IEEE International Conference on Healthcare Informatics, Imaging and Systems Biology, HISB 2011, San Jose, CA, USA, July 26-29, 2011, 250–57. IEEE, 2011.
 150. [ICDM 11] Cao, Nan, David Gotz, **Jimeng Sun**, Yu-Ru Lin, and Huamin Qu. "SolarMap: Multifaceted Visual Analytics for Topic Exploration." ICDM 2011
 151. [AMIA 11] Gotz, David, **Jimeng Sun**, Nan Cao, and Shahram Ebadollahi. "Visual Cluster Analysis in Support of Clinical Decision Intelligence." AMIA 2011 (2011): 481–90.
 152. [SDM 11] Kang, U., Spiros Papadimitriou, **Jimeng Sun**, and Hanghang Tong. "Centralities in Large Networks: Algorithms and Observations." SDM 2011.
 153. [KDD 11] Kang, U., Hanghang Tong, **Jimeng Sun**, Ching-Yung Lin, and Christos Faloutsos. "GBASE: A Scalable and General Graph Management System." KDD 2011.
 154. [ASONAM 11] Wang, Chi, Jie Tang, **Jimeng Sun**, and Jiawei Han. "Dynamic Social Influence Analysis through Time-Dependent Factor Graphs." In International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2011,
 155. [AAAI 11] Wang, Fei, Noah Lee, **Jimeng Sun**, Jianying Hu, and Shahram Ebadollahi. "Automatic Group Sparse Coding." AAAI 2011.
 156. [TVCG 11] Nan Cao, David Gotz, **Jimeng Sun**, and Huamin Qu. "DICON: Interactive Visual Analysis of Multidimensional Clusters." IEEE Transactions on Visualization and Computer Graphics 17, no. 12 (December 2011): 2581–90. doi:10.1109/TVCG.2011.188. (impact factor 1.4)
 157. [TKDD 11] Lin, Yu-Ru, **Jimeng Sun**, Hari Sundaram, Aisling Kelliher, Paul Castro, and Ravi B. Konuru. "Community Discovery via Metagraph Factorization." TKDD 5, no. 3 (2011).
 158. [SDM 11] Wang, Fei, **Jimeng Sun**, and Shahram Ebadollahi. "Integrating Distance Metrics Learned from Multiple Experts and Its Application in Inter-Patient Similarity Assessment." SDM 2011.
 159. [SDM 11] Wang, Fei, **Jimeng Sun**, Jianying Hu, and Shahram Ebadollahi. "iMet: Interactive Metric Learning in Healthcare Applications." SDM 2011.
 160. [AMIA 10] Ebadollahi, Shahram, **Jimeng Sun**, David Gotz, Jianying Hu, Daby Sow, and Chalapathy Neti. "Predicting Patient's Trajectory of Physiological Data Using Temporal Trends in Similar Patients: A System for Near-Term Prognostics." AMIA 2010 (2010): 192–96.
 161. [SDM 10] Lin, Yu-Ru, **Jimeng Sun**, Nan Cao, and Shixia Liu. "ContexTour: Contextual Contour Analysis on Dynamic Multi-Relational Clustering." SDM 2010.
 162. [ICDM 10] **Jimeng Sun**, Daby M. Sow, Jianying Hu, and Shahram Ebadollahi. "A System for Mining Temporal Physiological Data Streams for Advanced Prognostic Decision Support." ICDM 2010.
 163. [KDD 10] Tan, Chenhao, Jie Tang, **Jimeng Sun**, Quan Lin, and Fengjiao Wang. "Social Action Tracking via Noise Tolerant Time-Varying Factor Graphs." KDD 2010.
 164. [CIKM 10] Xiang, Liang, Quan Yuan, Shiwan Zhao, Li Chen, Xiatian Zhang, Qing Yang, and **Jimeng Sun**. "Temporal Recommendation on Graphs via Long- and Short-Term Preference Fusion." KDD 2010.
 165. [CIKM 10] Zhang, Duo, **Jimeng Sun**, ChengXiang Zhai, Abhijit Bose, and Nikos Anerousis. "PTM: Probabilistic Topic Mapping Model for Mining Parallel Document Collections." In Proceedings of the 19th ACM Conference on Information and Knowledge Management, CIKM 2010.

-
166. [ICDM 10] Zhang, Yuan, Jie Tang, **Jimeng Sun**, Yiran Chen, and Jinghai Rao. "MoodCast: Emotion Prediction via Dynamic Continuous Factor Graph Model." ICDM 2010.
 167. [TVCG 10] Nan Cao, **Jimeng Sun**, Yu-Ru Lin, David Gotz, Shixia Liu, and Huamin Qu. "FacetAtlas: Multifaceted Visualization for Rich Text Corpora." IEEE Transactions on Visualization and Computer Graphics 16, no. 6 (December 2010): 1172–81. doi:10.1109/TVCG.2010.154. (impact factor 1.4)
 168. [EMBSC 10] Sow, Daby, Alain Biem, **Jimeng Sun**, Jianying Hu, and Shahram Ebadollahi. "Real-Time Prognosis of ICU Physiological Data Streams." Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2010 (2010): 6785–88. doi:10.1109/IEMBS.2010.5625983.
 169. [EMBSC 10] **Jimeng Sun**, David Gotz, and Nan Cao. "DiseaseAtlas: Multi-Facet Visual Analytics for Online Disease Articles." Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference 2010 (2010): 1123–26. doi:10.1109/IEMBS.2010.5627103.
 170. [ICPR 10] **Jimeng Sun**, Daby M. Sow, Jianying Hu, and Shahram Ebadollahi. "Localized Supervised Metric Learning on Temporal Physiological Data." In 20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010, 4149–52. IEEE, 2010. doi:10.1109/ICPR.2010.1009.
 171. [ICDE 09] Lin, Ching-Yung, Nan Cao, Shixia Liu, Spiros Papadimitriou, **Jimeng Sun**, and Xifeng Yan. "SmallBlue: Social Network Analysis for Expertise Search and Collective Intelligence." In Proceedings of the 25th International Conference on Data Engineering, ICDE 2009, March 29 2009 - April 2 2009, Shanghai, China, 1483–86. IEEE, 2009.
 172. [PSVP 09] Shi, Lei, Nan Cao, Shixia Liu, Weihong Qian, Li Tan, Guodong Wang, **Jimeng Sun**, and Ching-Yung Lin. "HiMap: Adaptive Visualization of Large-Scale Online Social Networks." In IEEE Pacific Visualization Symposium PacificVis 2009, Beijing, China, April 20-23, 2009, 41–48. IEEE Computer Society, 2009.
 173. [WWW 09] Lin, Yu-Ru, **Jimeng Sun**, Paul Castro, Ravi B. Konuru, Hari Sundaram, and Aisling Kelliher. "Extracting Community Structure through Relational Hypergraphs." In Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009,
 174. [KDD 09] Lin, Yu-Ru, **Jimeng Sun**, Paul Castro, Ravi B. Konuru, Hari Sundaram, and Aisling Kelliher. "MetaFac: Community Discovery via Relational Hypergraph Factorization." KDD 2009.
 175. [SDM 09] **Jimeng Sun**, Spiros Papadimitriou, Ching-Yung Lin, Nan Cao, Shixia Liu, and Weihong Qian. "MultiVis: Content-Based Social Network Exploration through Multi-Way Visual Analysis." SDM 2009.
 176. [KDD 09] Tang, Jie, **Jimeng Sun**, Chi Wang, and Zi Yang. "Social Influence Analysis in Large-Scale Networks." KDD 2009.
 177. [ICDM 09] Wang, Fei, **Jimeng Sun**, Tao Li, and Nikos Anerousis. "Two Heads Better Than One: Metric+Active Learning and Its Applications for IT Service Classification." ICDM 2009.
 178. [IBM JRD 09] Khan, Asheq, Hani Jamjoom, and **Jimeng Sun**. "AIM-HI: A Framework for Request Routing in Large-Scale IT Global Service Delivery." IBM Journal of Research and Development 53, no. 6 (2009): 4. doi:10.1147/JRD.2009.5429032.
 179. [SIGKDD Explorations 08] **Jimeng Sun**. "Incremental Pattern Discovery on Streams, Graphs and Tensors." SIGKDD Explorations 10, no. 2 (2008): 28–29.
 180. [TKDD 08] **Jimeng Sun**, Dacheng Tao, Spiros Papadimitriou, Philip S. Yu, and Christos Faloutsos. "Incremental Tensor Analysis: Theory and Applications." TKDD 2, no. 3 (2008).
 181. [DMKD 08] **Jimeng Sun**, Charalampos E. Tsourakakis, Evan Hoke, Christos Faloutsos, and Tina Eliassi-Rad. "Two Heads Better than One: Pattern Discovery in Time-Evolving Multi-Aspect Data." Data Min. Knowl. Discov. 17, no. 1 (2008): 111–28. doi:10.1007/s10618-008-0112-3. (impact factor 2.714)

182. [SADM 08] **Jimeng Sun**, Yinglian Xie, Hui Zhang, and Christos Faloutsos. "Less Is More: Sparse Graph Mining with Compact Matrix Decomposition." *Statistical Analysis and Data Mining* 1, no. 1 (2008): 6–22. doi:10.1002/sam.102.
183. [IEEE TCSVT 08] Tao, Dacheng, Mingli Song, Xuelong Li, Jialie Shen, **Jimeng Sun**, Xindong Wu, Christos Faloutsos, and Stephen J. Maybank. "Bayesian Tensor Approach for 3-D Face Modeling." *IEEE Trans. Circuits Syst. Video Techn.* 18, no. 10 (2008): 1397–1410.
184. [WCCI 08] Tao, Dacheng, **Jimeng Sun**, Jialie Shen, Xindong Wu, Xuelong Li, Stephen J. Maybank, and Christos Faloutsos. "Bayesian Tensor Analysis." In *Proceedings of the International Joint Conference on Neural Networks, IJCNN 2008, Part of the IEEE World Congress on Computational Intelligence, WCCI 2008, Hong Kong, China, June 1-6, 2008*.
185. [CIKM 08] Boutsidis, Christos, **Jimeng Sun**, and Nikos Anerousis. "Clustered Subset Selection and Its Applications on IT Service Metrics." In *Proceedings of the 17th ACM Conference on Information and Knowledge Management, CIKM 2008*.
186. [ICDM 08] Kolda, Tamara G., and **Jimeng Sun**. "Scalable Tensor Decompositions for Multi-Aspect Data Mining." *ICDM 2008*. (Best research paper)
187. [ICDM 08] Papadimitriou, Spiros, and **Jimeng Sun**. "DisCo: Distributed Co-Clustering with Map-Reduce: A Case Study towards Petabyte-Scale End-to-End Mining." *ICDM 2008*.
188. [PKDD 08] Papadimitriou, Spiros, **Jimeng Sun**, Christos Faloutsos, and Philip S. Yu. "Hierarchical, Parameter-Free Community Discovery." In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML/PKDD 2008*.
189. [PKDD 08] **Jimeng Sun**, Charalampos E. Tsourakakis, Evan Hoke, Christos Faloutsos, and Tina Eliassi-Rad. "Two Heads Better Than One: Pattern Discovery in Time-Evolving Multi-Aspect Data." In *Machine Learning and Knowledge Discovery in Databases, European Conference, ECML/PKDD 2008*.
190. [KDD 08] Tong, Hanghang, Spiros Papadimitriou, **Jimeng Sun**, Philip S. Yu, and Christos Faloutsos. "Colibri: Fast Mining of Large Static and Dynamic Graphs." *KDD 2008*.
191. [ICDE 07] Li, Feifei, **Jimeng Sun**, Spiros Papadimitriou, George A. Mihaila, and Ioana Stanoi. "Hiding in the Crowd: Privacy Preservation on Evolving Streams through Correlation Tracking." In *Proceedings of the 23rd International Conference on Data Engineering, ICDE 2007, The Marmara Hotel, Istanbul, Turkey, April 15-20, 2007*, 686–95. IEEE, 2007.
192. [KDD 07] **Jimeng Sun**, Christos Faloutsos, Spiros Papadimitriou, and Philip S. Yu. "GraphScope: Parameter-Free Mining of Large Time-Evolving Graphs." *KDD 2007*.
193. [SDM 07] **Jimeng Sun**, Yinglian Xie, Hui Zhang, and Christos Faloutsos. "Less Is More: Compact Matrix Decomposition for Large Sparse Graphs." *SDM 2007*. (Best research paper)
194. [IJCAI 07] Xing Wei, **Jimeng Sun**, and Xuerui Wang. "Dynamic Mixture Models for Multiple Time-Series." In *IJCAI 2007, Proceedings of the 20th International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007*, 2909–14, 2007.
195. [ICONIP 2007] Tao, Dacheng, **Jimeng Sun**, Xindong Wu, Xuelong Li, Jialie Shen, Stephen J. Maybank, and Christos Faloutsos. "Probabilistic Tensor Analysis with Akaike and Bayesian Information Criteria." In *Neural Information Processing, 14th International Conference, ICONIP 2007, Kitakyushu, Japan, November 13-16, 2007*, 4984:791–801. *Lecture Notes in Computer Science*. Springer, 2007. doi:10.1007/978-3-540-69158-7_82.
196. [VLDB 06] Hoke, Evan, **Jimeng Sun**, and Christos Faloutsos. "InteMon: Intelligent System Monitoring on Large Clusters." In *Proceedings of the 32nd International Conference on Very Large Data Bases, Seoul, Korea, September 12-15, 2006*, 1239–42. ACM, 2006.
197. [PAKDD 06] **Jimeng Sun**, Spiros Papadimitriou, and Christos Faloutsos. "Distributed Pattern Discovery in Multiple Streams." In *Advances in Knowledge Discovery and Data Mining, 10th Pacific-Asia Conference, PAKDD 2006*.

-
198. [ICDM 06] Papadimitriou, Spiros, **Jimeng Sun**, and Philip S. Yu. “Local Correlation Tracking in Time Series.” In Proceedings of the 6th IEEE International Conference on Data Mining ICDM 2006.
 199. [ICDM 06] **Jimeng Sun**, Spiros Papadimitriou, and Philip S. Yu. “Window-Based Tensor Analysis on High-Dimensional and Multi-Aspect Streams.” In Proceedings of the 6th IEEE International Conference on Data Mining (ICDM) 2006.
 200. [KDD 06] **Jimeng Sun**, Dacheng Tao, and Christos Faloutsos. “Beyond Streams and Graphs: Dynamic Tensor Analysis.” KDD 2006.
 201. [OSR 06] Hoke, Evan, **Jimeng Sun**, John D. Strunk, Gregory R. Ganger, and Christos Faloutsos. “InteMon: Continuous Mining of Sensor Data in Large-Scale Self-Infrastructures.” Operating Systems Review 40, no. 3 (2006): 38–44.
 202. [IS 06] **Jimeng Sun**, Yufei Tao, Dimitris Papadias, and George Kollios. “Spatio-Temporal Join Selectivity.” Inf. Syst. 31, no. 8 (2006): 793–813. doi:10.1016/j.is.2005.02.002.
 203. [VLDB 05] Papadimitriou, Spiros, **Jimeng Sun**, and Christos Faloutsos. “Streaming Pattern Discovery in Multiple Time-Series.” In Proceedings of the 31st International Conference on Very Large Data Bases (VLDB), Trondheim, Norway, 2005.
 204. [ICDE 05] **Jimeng Sun**, Spiros Papadimitriou, and Christos Faloutsos. “Online Latent Variable Detection in Sensor Networks.” In Proceedings of the 21st International Conference on Data Engineering, ICDE 2005, 5-8 April 2005, Tokyo, Japan, 1126–27. IEEE Computer Society, 2005.
 205. [ICDE 04] **Jimeng Sun**, Dimitris Papadias, Yufei Tao, and Bin Liu. “Querying about the Past, the Present, and the Future in Spatio-Temporal.” In Proceedings of the 20th International Conference on Data Engineering, ICDE 2004.
 206. [VLDB 03] Tao, Yufei, Dimitris Papadias, and **Jimeng Sun**. “The TPR*-Tree: An Optimized Spatio-Temporal Access Method for Predictive Queries.” In VLDB, 790–801, 2003.
 207. [ICDE 03] Tao, Yufei, **Jimeng Sun**, and Dimitris Papadias. “Selectivity Estimation for Predictive Spatio-Temporal Queries.” ICDE 2003.
 208. [TODS 03] Tao, Yufei, **Jimeng Sun**, and Dimitris Papadias. “Analysis of Predictive Spatio-Temporal Queries.” ACM Trans. Database Syst. (TODS) 28, no. 4 (2003): 295–336.
 209. [IEEE DEB 02] Papadias, Dimitris, Yufei Tao, Jun Zhang, Nikos Mamoulis, Qiongmao Shen, and **Jimeng Sun**. “Indexing and Retrieval of Historical Aggregate Information about Moving Objects.” IEEE Data Eng. Bull. 25, no. 2 (2002): 10–17.

PATENT

1. Cao, Nan, David H. Gotz, and **Jimeng Sun**. Visual analysis of multidimensional clusters. USPTO 9342579. US Patent, filed May 31, 2011, and issued May 17, 2016.
2. Papadimitriou, Spyridon, **Jimeng Sun**, and Philip S. Yu. 2009. Systems and methods for simultaneous summarization of data cube streams. USPTO 7505876. issued March 17, 2009.
3. Chang, Yuan-Chi, Feifei Li, Spyridon Papadimitriou, George A. Mihaila, Ioana Stanoi, **Jimeng Sun**, and Philip S. Yu. 2010. Preserving privacy of one-dimensional data streams using dynamic correlations. USPTO 7853545. US Patent, filed February 26, 2007, and issued December 14, 2010. <https://www.google.com/patents/US7853545>.
4. Lin, Ching-Yung, Spyridon Papadimitriou, **Jimeng Sun**, and Kun-Lung Wu. 2012. Content-based and time-evolving social network analysis. USPTO 8204988. US Patent, filed September 2, 2009, and issued June 19, 2012. <https://www.google.com/patents/US8204988>.
5. Chang, Yuan-Chi, Feifei Li, Spyridon Papadimitriou, George A. Mihaila, Ioana Stanoi, **Jimeng Sun**, and Philip S. Yu. 2010. Preserving privacy of one-dimensional data streams by perturbing data with noise and using dynamic autocorrelation. USPTO 7840516. US Patent, filed February 26, 2007, and issued November 23, 2010. <https://www.google.com/patents/US7840516>.

-
6. Ebadollahi, Shahram, Jianying Hu, Martin S. Kohn, Noah Lee, Robert K. Sorrentino, **Jimeng Sun**, and Fei Wang. 2013. Mining temporal patterns in longitudinal event data using discrete event matrices and sparse coding. USPTO 8583586. US Patent, filed January 21, 2011, and issued November 12, 2013. <https://www.google.com/patents/US8583586>.

KEYNOTE AND INVITED TALKS

1. Keynote “Deep learning for drug development” at KDD’20 workshop on Deep Learning on Graphs
2. Keynote at NIPS’16 workshop on learning with tensors
3. Keynote at SDM’13 Data Mining for Medicine and Healthcare “Large-scale Patient Similarity Learning for Healthcare Analytics”
4. Keynote at ORNL Biomedical Science/Engg. Center (BSEC) Annual conference 2013 “Large-scale Patient Similarity Learning for Healthcare Analytics”
5. Invited talk at Mayo Clinic 2020, “Deep learning for Drug Development”
6. Invited talk at Chinese University of Hong Kong July 2017, “Doctor AI: Interpretable Deep Learning Models for Healthcare Applications”
7. Invited talk at UMass Boston Sep 2016, “Doctor AI: Interpretable Deep Learning Models for Healthcare Applications”
8. Invited talk at Mass General Hospital July 2016, “Scalable Predictive Modeling Methods for Healthcare Applications”
9. Invited talk at Modern Massive Data Sets (MMDS) 2016 “Scalable Predictive Modeling Methods for Healthcare Applications”
10. Invited talk at University of Kentucky Jun 2016, “Scalable Predictive Modeling Methods for Healthcare Applications”
11. Invited talk at seminar series at Emory university, Apr 2016 “Computational Phenotyping using Tensor Factorization and Tensor Network”
12. Invited talk at Distinguished Seminar Series at Florida International University, Nov 2015 “Building Scalable Health Analytic Platform: Computational Phenotyping and Cloud-based Predictive Modeling”
13. Contributed Talk at Machine learning healthcare conference 2016 “Doctor AI: Predicting Clinical Events via Recurrent Neural Networks”
14. Invited talk at Rutgers university Oct 2015, “Computational Phenotyping using Tensor Factorization”
15. Invited talk at CMU Sep 2015, “Computational Phenotyping using Tensor Factorization and Tensor Network
16. Invited talk, MUCMD, August 2015: “Computational Phenotyping using Knowledge Guided Tensor Factorization and Completion
17. Invited talk, Garvan university Australia, Aug 2015 “Predictive Modeling with Computational Phenotyping and Parallel Computing”
18. Invited talk UCSD Jun 2015, “Building a Scalable Health Analytic Platform”
19. Invited talk at Bioinformatics departmental seminar at UNCC Apr 2015 “Building a Scalable Predictive Modeling Platform for Healthcare Applications”
20. Invited talk at USC Apr 2015 “Building a Scalable Predictive Modeling Platform for Healthcare Applications”
21. Invited talk at MIT CSAIL Oct 2014 “Computational phenotyping through tensor factorization”
22. Invited talk at UCLA Jul 2014 “Do it once, Do it right - Building a Scalable Predictive Modeling Platform for Healthcare Applications”
23. AMIA-CRI 2016 “Advanced Machine Learning for Healthcare”

-
24. AMIA-CRI 2015 “Computational phenotyping”
 25. Biomedical and Healthcare Analytics on Big Data. AMIA 2013

STUDENT ADVISED

Ph.D. students

UIUC

1. Junyi Gao, Ph.D. student in Computer Science, Fall 2020 – present
2. Zhenbang Wu, Ph.D. student in Computer Science, Fall 2020 – present
3. Zhen Lin, Ph.D. student in Computer Science, Fall 2020 – present
4. Brandon Theodorou, Ph.D. student in Computer Science, Fall 2021 – present
5. Chaoqi Yang, Ph.D. student in Computer Science, Fall 2021 – present
6. Zifeng Wang, Ph.D. student in Computer Science, Fall 2021 – present

GaTech

7. Siddharth Biswal, Ph.D. candidate in Computer Science, 2016 – present
8. Tianfan Fu, Ph.D. candidate in Computer Science, 2018 - present
9. Kunal Malhotra, graduated with Ph.D. in Computer Science, 2016, current position: Principal Data Scientist at Anthem
10. Robert Chen, 2014 – May 2018, MD-PhD Emory/GT (defended in May 2018)
11. Ioakeim Perros, graduated with Ph.D. in Computer Science, 2019, current position: Lead Machine Learning Scientist at HEALTH[at]SCALE
12. Edward Choi, 2014 – graduated with Ph.D. in Computer Science, 2018, current position: Assistant Professor at KAIST Graduate School of AI

IQVIA

1. Junyuan Shang, MS student in Computer Science at Peking University, 2018 - 19, current position: machine learning scientist at Baidu NLP
2. Xinyao Zhang, MS student in Computer Science at Tsinghua University, 2019, current position: machine learning scientist at Tencent AI Lab
3. Kexin Huang, BS student in Computer Science at NYU, 2019 - 21, current position: Ph.D. student in Computer Science at Stanford University
4. Limeng Cui, PhD student in Computer Science at PSU, Summer 2019
5. Chacha Chen, PhD student in Computer Science at PSU, Summer 2020
6. Rakshith Sharma Srinivasa, PhD candidate in EE at GaTech, Summer 2020, current position: Senior Researcher at Samsung Research America
7. Nikos Kargas, PhD candidate in EE at UMN, Summer 2020, current position: Applied Scientist at Amazon TTS Research

IBM Research

13. Joyce Ho, University of Texas at Austin, 2013, Unsupervised Phenotyping using Tensor Factorization, current position: Assistant Professor at Emory University
14. Jiayu Zhou, Arizona State University, 2012, Stable Feature Selection and its applications in healthcare analytics, current position: Associate Professor at Michigan State University
15. Parikshit Sondhi, University of Illinois at Urbana-Champaign, 2011, Text mining on Clinical data, current position: Engineering Manager at Snap Inc.
16. Dijun Luo, University of Texas at Arlington, 2011, Feature Selection and predictive modeling on clinical data, current position: Data Scientist at RetailMeNot, Inc

-
17. Lei Li, Carnegie Mellon University, 2010, Time series analysis on ECG data, current position: Assistant Professor at University of California Santa Barbara
 18. Nan Cao, Hong Kong University of Science & Technology, 2010, Visualization on healthcare data, current position: Professor at Tongji University
 19. U Kang, Carnegie Mellon University, 2010, Graph mining on big data using Hadoop, current position: Associate Professor at Seoul National University.
 20. Duo Zhang, University of Illinois at Urbana-Champaign, 2009, Text mining on IT support data, current position: Engineering Manager at Pinterest
 21. Yu-Ru Lin, Arizona State University, 2008, Social Media Analytics using Tensor analysis, current position: Associate Professor at University of Pittsburgh
 22. Christos Boutsidis, Rensselaer Polytechnic Institute, 2008, Column selection algorithm and its application to IT support data, current position: Head of Engineering at Goldman Sachs

M.S. and undergraduate students

23. Chaoqi Yang, 2021, current position: Ph.D. student in CS at UIUC
24. Olivier Deiss, 2018, current position: Software Engineer at Google
25. James Mullenbach, 2018, current position: Software Engineer at Google
26. Apurv Verma, 2016, current position: Applied Scientist at Amazon
27. Aashu, Singh, 2015, current position: Software Engineer (Machine Learning) at Facebook
28. Joshua Kulas, 2015-17, current position: Software Development Engineer at Amazon
29. Keegan Nesbitt, 2015, current position: AI/ML Engineer at GSK
30. Sizhe Lin, 2015, current position: Software Engineer at YouTube

PostDoc and Other Mentorship

31. Jiajia Li, 2019, current position: Assistant Professor at William and Mary University
32. Shenda Hong, 2019 - 2020, current position: Postdoctoral Researcher in National Institute of Health Data Science at Peking University
33. Taha Bahadori, 2015 – 2017, current position: Senior Machine Learning Scientist at Amazon
34. Mohammed Khalilia, 2014 – 2015, current position: Machine Learning Scientist at Amazon

PROFESSIONAL SERVICE

[Organizing Committee] KDD (2011, 2012, 2013), SDM (2012)

[Associate Editor] Associated editor for Transaction of Knowledge Discovery from Data (TKDD)

[Program Committee Chair] CIKM (2017), AIME (2015), AIME'15 workshop on Matrix Computations for Biomedical Informatics, KDD'14 workshop on health informatics (HI-KDD), KDD'13 workshop on data mining for healthcare, ICML'13 workshop on Role of Machine Learning in Transforming Healthcare, KDD'11 workshops: Large-scale Data Mining: Theory and Applications, KDD'11 workshops: Data Mining for Medicine and Healthcare; VisWeek'10 workshop: Visual Analytics in Healthcare; KDD'10 workshop: LDM: Large-scale Data Mining: theory and applications; ICDM'09 workshop: Large-scale Data Mining: theory and applications

[Senior Program Committee] AAAI (2021, 2020), KDD (2022, 2020, 2018, 2017, 2016, 2015, 2014, 2013), IJCAI (2020), SDM (2013, 2011, 2010), CIKM (2016, 2015)

[Program Committee] KDD (2012, 2011, 2010, 2009, 2008, 2007), NeurIPS (2021, 2020, 2019, 2018, 2017, 2016), ICLR (2021, 2020, 2019), ICML (2021, 2020), CIKM (2014, 2012, 2011, 2010, 2009, 2008), ICDM (2016, 2012, 2011, 2010, 2009)