Yanshan Wang, PhD, FAMIA

CONTACT INFORMATION	Department of Health Information Management
	University of Pittsburgh yanshan.wang@pitt.edu
	6026 Forbes Tower, Pittsburgh, PA 15260

RESEARCH INTERESTS

My research focuses on artificial intelligence, natural language processing, machine/deep learning methodologies and applications in healthcare.

EDUCATION

Mayo Clinic, Rochester, USA

Postdoc, Biomedical Informatics, Mar 2018

- Research Topic: *Clinical Natural Language Processing and Biomedical Informatics.*
- Advisors: Hongfang Liu, Ph.D

Korea University, Seoul, Korea

Ph.D., Information Management Engineering, Feb 2015

- GPA: 4.25/4.5
- Thesis: *A Study of Ensemble Models for Document Ranking in Information Retrieval.*
- Advisors: In-Chan Choi, Ph.D

M.S., Information Management Engineering, Aug 2012

- GPA: 4.0/4.5
- Thesis: *Stock price direction prediction using principal component analysis and support vector machine.*
- Advisors: In-Chan Choi, Ph.D

Harbin Institute of Technology, Harbin, China

B.S., Computer Science and Technology, July 2010

- GPA: 85/100
- *Summa Cum Laude*

EXPERIENCE

Assistant Professor of Health Informatics	Jun 2021 to present
Department of Health Information Management, University of Pittsburgh	
Vice Chair of Research	Jun 2021 to present
Department of Health Information Management, University of Pittsburgh	
Assistant Professor of Intelligent Systems	Jan 2022 to present
Intelligent Systems Program, University of Pittsburgh	
Chair-Elect	Jan 2022 to present
Natural Language Processing (NLP) Working Group, American Medical Informatics Association (AMIA)	
Assistant Professor of Biomedical Informatics	Oct 2019 to present

Division of Digital Health Sciences,
Department of Health Sciences Research,
Mayo Clinic

Associate Consultant I   Oct 2019 to May 2021
Division of Digital Health Sciences,
Department of Health Sciences Research,
Mayo Clinic

Research Associate   Mar 2018 to Oct 2019
Department of Health Sciences Research,
Mayo Clinic

Research Fellow   Mar 2015 to Mar 2018
Department of Health Sciences Research,
Mayo Clinic

Research Assistant   Sep 2010 to Feb 2015
System Optimization Lab,
Korea University

Assistant Engineer   Oct 2005 to July 2006
Artificial Intelligence and Bioinformatics Lab,
Harbin Institute of Technology

HONORS	Fellow of American Medical Informatics Association (FAMIA)	2020
	Early Career Reviewer, National Institute of Health	2021

PROFESSIONAL	Member, American Medical Informatics Association (AMIA)	2015-
MEMBERSHIPS	Member, Association for Computing Machinery (ACM)	2016-
	Member, Association for Computational Linguistics (ACL),	2016-

PUBLICATIONS

1. Nicolas Nunez, Joanna M Biernacka, Manuel Gardea-Resendez, Bhavani Singh Agnikula Kshatriya, Euijung Ryu, Sunyang Fu, Balwinder Singh, Brandon Coombes, Mark Frye, **Yanshan Wang**. "Natural Language Processing for Automatic Identification of Major Depressive Disorders in Free-Text Electronic Health Records". Biological Psychiatry 89, no. 9 (2021): S155.

2. Coombes B, Landi I, Jenkins G, Ryu E, Nunez N, **Wang Y**, Colby C, Batzler A, Weissman M, Olfson M, Wickramaratne P. POLYGENIC RISK OF COMORBID ANXIETY AND DEPRESSION ACROSS HEALTH SYSTEMS. European Neuro psychopharmacology. 2021 Oct 1;51:e31-2.

3. Ryu, Euijung, Gregory Jenkins, **Yanshan Wang**, Mark Olfson, Ardesheer Talati, Lauren Lepow, Alexander Charney et al. "Relative Importance of Self-Reported Social Determinants of Health for Risk of Depression." Biological Psychiatry 89, no. 9 (2021): S97.

4. Bhavani Singh Agnikula Kshatriya, Nicolas A. Nunez, Manuel Gardea Resendez, Euijung Ryu, Brandon J. Coombes, Sunyang Fu, Mark A. Frye, Joanna M. Biernacka, and **Yanshan Wang**. "Neural Language Models with Distant Supervision to Identify Major Depressive Disorder from Clinical Notes." arXiv preprint arXiv: 2104.09644. 2021.

5. Euijung Ryu, Gregory Jenkins, **Yanshan Wang**, Mark Olfson, Ardesheer Talati, Lauren Lepow, Alexander Charney et al. "Relative Importance of Self-Reported Social Determinants of Health for Risk of Depression." Biological Psychiatry 89, no. 9 (2021): S97.

6. Marika Cusick, Prakash Adekkanattu, Thomas R. Campion Jr, Evan T. Sholle, Annie Myers, Samprit Banerjee, George Alexopoulos, **Yanshan Wang**, and Jyotishman Pathak. "Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation." Journal of psychiatric research. 2021.

7. Anusha Bompelli*, **Yanshan Wang***, Ruyuan Wan, Esha Singh, Yuqi Zhou, Lin Xu, David Oniani, Bhavani Singh Agnikula Kshatriya, E. Balls-Berry, and Rui Zhang. "Social determinants of health in the era of artificial intelligence with electronic health records: A scoping review." Health Data Science (2021). [*co-first authors]

8. Zitao Shen, Yoonkwon Yi, Anusha Bompelli, Fang Yu, **Yanshan Wang**, and Rui Zhang. "Extracting Lifestyle Factors for Alzheimer's Disease from Clinical Notes Using Deep Learning with Weak Supervision." arXiv preprint arXiv:2101.09244. 2021.

9. Marika Cusick, Prakash Adekkanattu, Thomas R. Campion Jr, Evan T. Sholle, Annie Myers, Samprit Banerjee, George Alexopoulos, **Yanshan Wang**, and Jyotishman Pathak. "Using Weak Supervision and Deep Learning to Classify Clinical Notes for Identification of Current Suicidal Ideation." Journal of Psychiatric Research. 2021.

10. Feichen Shen, Sijia Liu, Sunyang Fu, **Yanshan Wang**, Sam Henry, Ozlem Uzuner, and Hongfang Liu. "Family History Extraction From Synthetic Clinical Narratives Using Natural Language Processing: Overview and Evaluation of a Challenge Data Set and Solutions for the 2019 National NLP Clinical Challenges (n2c2)/Open Health Natural Language Processing (OHNLP) Competition." JMIR Medical Informatics 9, no. 1 (2021): e24008.

11. **Yanshan Wang**, Sunyang Fu, Feichen Shen, Sam Henry, Ozlem Uzuner, and Hongfang Liu. The 2019 n2c2/OHNLP Track on Clinical Semantic Textual Similarity: Overview. JMIR Medical Informatics. 2020.

12. Bhavani Agnikula Kshatriya, Joy Balls-Berry, **Yanshan Wang**. Completeness of Social Determinants of Health in Electronic Health Records: A case study on the Patient-Provided Information from a minority cohort with sexually transmitted diseases. under review.

13. Andrew Wen, Liwei Wang, Huan He, Sijia Liu, Sunyang Fu, Sunghwan Sohn, Jacob A.Kugel, Vinod C. Kaggal, Ming Huang, **Yanshan Wang**, Feichen Shen, Jungwei Fan, Hongfang Liu. "An aberration detection-based approach for sentinel syndromic surveillance of covid-19 and other novel influenza-like illnesses." Journal of Biomedical Informatics 113 (2021): 103660.

14. David Oniani, **Yanshan Wang**. A Qualitative Evaluation of Language Models on Automatic Question-Answering for COVID-19. *the 2020 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM-BCB)*. 2020.

15. Sijia Liu, **Yanshan Wang**, Andrew Wen, Liwei Wang, Na Hong, Feichen Shen, Steven Bedrick, William Hersh, and Hongfang Liu. Implementation of a Cohort Retrieval System for Clinical Data Repositories Using the Observational Medical Outcomes Partnership Common Data Model: Proof-of-Concept System Validation. JMIR medical informatics, 2020.

16. Luke A. Carlson, Molly M. Jeffery, Sunyang Fu, Huan He, Rozalina G. McCoy, **Yanshan Wang**, W. Michael Hooten, Jennifer St Sauver, Hongfang Liu, and Jungwei Fan. Characterizing Chronic Pain Episodes in Clinical Text: Comprehensive Annotation and Corpus Analysis at Two Health Care Systems. JMIR Medical Informatics, 2020.

17. Steven R. Chamberlin, Steven D. Bedrick, Aaron M. Cohen, **Yanshan Wang**, Andrew Wen, Sijia Liu, Hongfang Liu, and William R. Hersh. Evaluation of patient-level retrieval from electronic health record data for a cohort discovery task. JAMIA Open, 2020.

18. Sam Henry, **Yanshan Wang**, Feichen Shen, and Ozlem Uzuner. "The 2019 National Natural language processing (NLP) Clinical Challenges (n2c2)/Open Health NLP (OHNLP) shared task on clinical concept normalization for clinical records." Journal of the American Medical Informatics Association (2020).

19. Nathan D Seligson, Jeremy L Warner, William S Dalton, David Martin, Robert S Miller, Debra Patt, Kenneth L Kehl, Matvey B Palchuk, Gil Alterovitz, Laura K Wiley, Ming Huang, Feichen Shen, **Yanshan Wang**, Khoa A Nguyen, Anthony F Wong, Funda Meric-Bernstam, Elmer V Bernstam, James L Chen, Recommendations for patient similarity classes: results of the AMIA 2019 workshop on defining patient similarity, Journal of the American Medical Informatics Association.

20. Sunyang Fu, David Chen, Huan He, Sijia Liu, Sungrim Moon, Kevin J Peterson, Feichen Shen, Liwei Wang, **Yanshan Wang**, Andrew Wen, Yiqing Zhao, Sunghwan Sohn, Hongfang Liu. Clinical Concept Extraction: a Methodology Review. *Journal of Biomedical Informatics.* 2020.

21. Fu S, Carlson LA, Peterson KJ, Wang N, Zhou X, Peng S, Jiang J, **Wang Y**, Sauver JS, Liu H. Natural Language Processing for the Evaluation of Methodological Standards and Best Practices of EHR-based Clinical Research. AMIA Jt Summits Transl Sci Proc. 2020; 2020:171-180.

22. **Yanshan Wang**, Yiqing Zhao, Terry M. Therneau, Elizabeth J. Atkinson, Ahmad P. Tafti, Nan Zhang, Shreyasee Amin, Andrew H. Limper, Sundeep Khosla, and Hongfang Liu. Unsupervised Machine Learning for the Discovery of Latent Disease Clusters and Patient Subgroups Using Electronic Health Records. *Journal of Biomedical Informatics.* 2020.

23. Andrew Wen, **Yanshan Wang**, Vinod C. Kaggal, Sijia Liu, Hongfang Liu, and Jungwei Fan. Enhancing Clinical Information Retrieval through Context-Aware Queries and Indices. *the 2019 IEEE International Conference on Big Data (Big Data)*, 2019.

24. Xin Zhou, **Yanshan Wang***, Sunghwan Sohn, Terry M. Therneau, Hongfang Liu, and David S. Knopman. Automatic Extraction and Assessment of Lifestyle Exposures for Alzheimer's Disease using Natural Language Processing. *International Journal of Medical Informatics.* 2019. [*corresponding author]

25. Ahmad P Tafti, **Yanshan Wang**, Feichen Shen, Elham Sagheb, Paul Kingsbury, Hongfang Liu. Integrating word embedding neural networks with PubMed abstracts to extract keyword proximity of chronic diseases. *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. 2019.

26. **Yanshan Wang**, Andrew Wen, Sijia Liu, William Hersh, Steven Bedrick, Hongfang Liu. Test collections for electronic health record-based clinical information retrieval. *JAMIA Open*. 2019.

27. **Yanshan Wang**, Hua Xu, and Ozlem Uzuner. The second international workshop on health natural language processing (HealthNLP 2019). *BMC Medical Informatics and Decision Making*. 2019: 1-3.

28. Liwei Wang, Lei Luo, **Yanshan Wang**, Jason Wampfler, Ping Yang, and Hongfang Liu. Natural language processing for populating lung cancer clinical research data. *BMC Medical Informatics and Decision Making* 19, no. 5 (2019): 239.

29. **Yanshan Wang**, Saeed Mehrabi, Sunghwan Sohn, Elizabeth J Atkinson, Shreyasee Amin, Hongfang Liu. Natural Language Processing of Radiology Reports for Identification of Skeletal Site-Specific Fractures. *BMC Medical Informatics and Decision Making*. 2019.

30. Feichen Shen, Suyuan Peng, Yadan Fan, Andrew Wen, Sijia Liu, **Yanshan Wang**, Liwei Wang, Hongfang Liu. HPO2Vec+: Leveraging Heterogeneous Knowledge Resources to Enrich Node Embeddings for the Human Phenotype Ontology. *Journal of biomedical informatics*. 2019.

31. Liwei Wang, **Yanshan Wang**, Feichen Shen, Majid Rastegar-Mojarad, and Hongfang Liu. Discovering associations between problem list and practice setting. *BMC medical informatics and decision making*. 2019.

32. Rezarta Islamaj Dogan, Sun Kim, Andrew Chatr-aryamontri, Chih-Hsuan Wei, Donald C Comeau, Rui Antunes, Sérgio Matos, Qingyu Chen, Aparna Elangovan, Nagesh C Panyam, Karin Verspoor, Hongfang Liu, **Yanshan Wang**, Zhuang Liu, Berna Altnel, Zehra Melce Hüsünbeyi, Arzucan Özgür, Aris Fergadis, Chen-Kai Wang, Hong-Jie Dai, Tung Tran, Ramakanth Kavuluru, Ling Luo, Albert Steppi, Jinfeng Zhang, Jinchan Qu, Zhiyong Lu. Overview of the BioCreative VI Precision Medicine Track: mining protein interactions and mutations for precision medicine. *Database*. 2019.

33. Vinod Vydiswaran, Yaoyun Zhang, **Yanshan Wang**, and Hua Xu. Special issue of BMC medical informatics and decision making on health natural language processing. *BMC Medical Informatics and Decision Making*. 2019.

34. Sungrim Moon, Sijia Liu, David Chen, **Yanshan Wang**, Douglas L Wood, Rajeev Chaudhry, Hongfang Liu, Paul Kingsbury. Salience of Medical Concepts of Inside Clinical Texts and Outside Medical Records for Referred Cardiovascular Patients. *Journal of Healthcare Informatics Research*. 2019.

35. **Yanshan Wang**, Sunghwan Sohn, Sijia Liu, Feichen Shen, Liwei Wang, Elizabeth J Atkinson, Shreyasee Amin, Hongfang Liu. A clinical text classification paradigm using weak supervision and deep representation. *BMC Medical Informatics and Decision Making*. 2019.

36. Feichen Shen, Yiqing Zhao, Liwei Wang, Majid Rastegar Mojarad, **Yanshan Wang**, Sijia Liu, Hongfang Liu. Rare Disease Knowledge Enrichment through a Data-Driven Approach. *BMC Medical Informatics and Decision Making.* 2019.

37. **Yanshan Wang**, Sijia Liu, Naveed Afzal, Majid Rastegar-Mojarad, Liwei Wang, Feichen Shen, Hongfang Liu. A Comparison of Word Embeddings for the Biomedical Natural Language Processing *Journal of Biomedical Informatics.* 2018.

38. **Yanshan Wang**, Liwei Wang, Majid Rastegar-Mojarad, Sungrim Moon, Feichen Shen, Naveed Afzal, Sijia Liu, Yuqun Zeng, Saeed Mehrabi, Sunghwan Sohn, Hongfang Liu. Clinical Information Extraction Applications: A Literature Review. *Journal of Biomedical Informatics*, 2018.

39. **Yanshan Wang**, Naveed Afzal, Sunyang Fu, Liwei Wang, Feichen Shen, Majid Rastegar-Mojarad, and Hongfang Liu. MedSTS: a resource for clinical semantic textual similarity. *Language Resources and Evaluation.* 2018.

40. Yiqing Zhao, Yanshan Wang, Henry Wang, Benjamin Yan, Feichen Shen, Kevin J Peterson, Walter A Rocca, Jennifer St Sauver, Hongfang Liu. Annotating Cohort Data Elements with OHDSI Common Data Model to Promote Research Reproducibility. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM).* 2018.

41. **Yanshan Wang**, Saeed Mehrabi, Sunghwan Sohn, Elizabeth Atkinson, Shreyasee Amin, Hongfang Liu. Automatic Extraction of Major Osteoporotic Fractures from Radiology Reports using Natural Language Processing. *IEEE International Conference on Healthcare Informatics Workshop.* 2018.

42. Majid Rastegar-Mojarad, Sijia Liu, **Yanshan Wang**, Naveed Afzal, Liwei Wang, Feichen Shen, Sunyang Fu, and Hongfang Liu. BioCreative/OHNLP Challenge 2018. *the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics.* 2018.

43. Feichen Shen, Sijia Liu, **Yanshan Wang**, Andrew Wen, Liwei Wang, and Hongfang Liu. Utilization of Electronic Medical Records and Biomedical Literature to Support the Diagnosis of Rare Diseases Using Data Fusion and Collaborative Filtering Approaches. *JMIR medical informatics.* 2018.

44. Sijia Liu, Feichen Shen, Ravikumar Komandur Elayavilli, **Yanshan Wang**, Majid Rastegar-Mojarad, Vipin Chaudhary, and Hongfang Liu. Extracting chemical-protein relations using attention-based neural networks. *Databas.* 2018.

45. Liwei Wang, Majid Rastegar-Mojarad, Zhiliang Ji, Sijia Liu, Ke Liu, Sungrim Moon, Feichen Shen, **Yanshan Wang**, Lixia Yao, John M Davis III, Hongfang Liu. Detecting pharmacovigilance signals combining electronic medical records with spontaneous reports: a case study of conventional disease-modifying antirheumatic drugs for rheumatoid arthritis. *Frontiers in Pharmacology.* 2018.

46. Xin Zhou, Hongfang Liu, **Yanshan Wang**. A Comparison of Lifestyle Interventions for Alzheimer's Disease Extracted from Clinical Notes and Literature. *IEEE International Conference on Healthcare Informatics Workshop.* 2018.

47. Feichen Shen, Sijia Liu, **Yanshan Wang**, Liwei Wang, Andrew Wen, Andrew H Limper, Hongfang Liu. Constructing Node Embeddings for Human Phenotype

Ontology to Assist Phenotypic Similarity Measurement. *IEEE International Conference on Healthcare Informatics Workshop.* 2018.

48. Susan McRoy, Majid Rastegar-Mojarad, **Yanshan Wang**, Kathryn J. Ruddy, Tufia C. Haddad, and Hongfang Liu. Assessing Unmet Information Needs of Breast Cancer Survivors: Exploratory Study of Online Health Forums Using Text Classification and Retrieval. *JMIR cancer.* 2018.

49. Liwei Wang, **Yanshan Wang**, Feichen Shen, Majid Rastegar-Mojarad, Hongfang Liu. Predicting Practice Setting Using Topic Modeling. *IEEE International Conference on Healthcare Informatics Workshop.* 2018.

50. Liwei Wang, Majid Rastegar-Mojarad, Ravikumar Komandur Elayavilli, **Yanshan Wang**, Hongfang Liu. Identification of Genetic Causality Statements in Medline Abstracts Leveraging Distant Supervision. *IEEE International Conference on Healthcare Informatics Workshop.* 2018.

51. **Yanshan Wang**, Naveed Afzal, Sijia Liu, Majid Rastegar-Mojarad, Liwei Wang, Feichen Shen, Sunyang Fu, and Hongfang Liu. Overview of the BioCreative/OHNLP Challenge 2018 Task 2: Clinical Semantic Textual Similarity. *Proceedings of the BioCreative/OHNLP Challenge.* 2018.

52. Sijia Liu, Majid Rastegar Mojarad, **Yanshan Wang**, Liwei Wang, Feichen Shen, Sunyang Fu, Hongfang Liu. Overview of the BioCreative/OHNLP 2018 Family History Extraction Task. *Proceedings of the BioCreative/OHNLP Challenge.* 2018.

53. **Yanshan Wang**, Majid Rastegar-Mojarad, Ravikumar Komandur-Elayavilli, Hongfang Liu. Leveraging word embeddings and medical entity extraction for biomedical dataset retrieval using unstructured texts. *Database*, 2017.

54. **Yanshan Wang**, Ravikumar Komandur-Elayavilli, Majid Rastegar-Mojarad, Hongfang Liu. Leveraging both Structured and Unstructured Data for Precision Information Retrieval. *Proceedings of the 26th Text REtrieval Conference (TREC)*, 2017.

55. Feichen Shen, Sijia Liu, **Yanshan Wang**, Liwei Wang, Naveed Afzal, Hongfang Liu. Leveraging Collaborative Filtering to Accelerate Rare Disease Diagnosis. *AMIA 2017 Annual Symposium*, 2017.

56. **Yanshan Wang**, Naveed Afzal, Liwei Wang, Feichen Shen, Hongfang Liu. Semantic Representations of Medical Terms: A Comparison Study of Word Embeddings Trained on Clinical Notes and PubMed Articles. *AMIA 2017 Annual Symposium*, 2017.

57. Naveed Afzal, Yanshan Wang, Feichen Shen, Liwei Wang, Hongfang Liu. A Resource for Clinical Semantic Textual Similarity. *AMIA 2017 Annual Symposium*, 2017.

58. Sungrim Moon, Sijia Liu, Paul Kingsbury, David Chen, **Yanshan Wang**, Feichen Shen, Rajeev Chaudhry, Hongfang Liu. Medical concept Intersection between outside medical records and consultant notes-a case study in transferred cardiovascular patients. *IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2017.

59. Majid Rastegar-Mojarad, Ravikumar Komandur Elayavilli, **Yanshan Wang**, Sijia Liu, Feichen Shen, Hongfang Liu. Semantic Information Retrieval: Exploring dependency and word embedding features in biomedical Information Retrieval. *Proceedings of BioCreative VI*, 2017.

60. **Yanshan Wang**, Feichen Shen, Ravikumar Komandur Elayavilli, Sijia Liu, Majid Rastegar-Mojarad, Hongfang Liu. Entity-enhanced Hierarchical Attention Neural Networks for Mining Protein Interactions from Biomedical Texts. *Proceedings of BioCreative VI*, 2017.

61. Sijia Liu, Feichen Shen, **Yanshan Wang**, Majid Rastegar-Mojarad, Ravikumar Komandur Elayavilli, Vipin Chaudhary, Hongfang Liu. Attention-based Neural Networks for Chemical Protein Relation Extraction. *Proceedings of BioCreative VI*, 2017.

62. Susan McRoy, Majid Rastegar-Mojarad, **Yanshan Wang**, Kathryn Ruddy, Tufia Haddad, Hongfang Liu. Content-based Analysis of Health Forum Text to Assess Potential the Unmet Needs of Breast Cancer Survivors. *Individualizing Medicine Conference*, 2017.

63. Sunghwan Sohn, **Yanshan Wang**, Chung-Il Wi, Elizabeth A Krusemark, Euijung Ryu, Mir H Ali, Young J Juhn, Hongfang Liu; Clinical documentation variations and NLP system portability: a case study in asthma birth cohorts across institutions. *Journal of the American Medical Informatics Association*, 2017.

64. Stephen Wu, Andrew Wen, **Yanshan Wang**, Sijia Liu, Hongfang Liu. Aligned-Layer Text Search in Clinical Notes. Studies in health technology and informatics, 2017.

65. Yuqun Zeng, Xusheng Liu, **Yanshan Wang**, Feichen Shen, Sijia Liu, Majid Rastegar-Mojarad, Liwei Wang, Hongfang Liu. Recommending Education Materials for Diabetic Questions Using Information Retrieval Approaches. *Journal of medical Internet research*, 2017.

66. Sijia Liu, **Yanshan Wang**, Na Hong, Feichen Shen, Stephen Wu, William Hersh, Hongfang Liu. On Mapping Textual Queries to a Common Data Model. *Proceedings of IEEE International Conference on Healthcare Informatics (ICHI)*, 2017.

67. **Yanshan Wang**, Sijia Liu, Majid Rastegar-Mojarad, Liwei Wang, Feichen Shen, Fei Liu, and Hongfang Liu. Dependency and AMR Embeddings for Drug-Drug Interaction Extraction from Biomedical Literature. *Proceedings of ACM-BCB'17*, 2017. DOI: 10.1145/3107411.3107426.

68. **Yanshan Wang**, Liwei Wang, Majid Rastegar-Mojarad, Sijia Liu, Feichen Shen, Hongfang Liu. Systematic Analysis of Free-Text Family History in Electronic Health Record *AMIA Summits on Translational Science Proceedings*, 2017.

69. Stephen Wu, Sijia Liu, **Yanshan Wang**, Tamara Timmons, Harsha Uppili, Steven Bedrick, William Hersh, Hongfang LiuIntra-institutional EHR Collections for Patient-Level Information Retrieval *Journal of the American Society for Information Science and Technology (JASIS&T)*, 2017.

70. **Yanshan Wang**, In-Chan Choi, Hongfang Liu. Generalized Ensemble Model for Document Ranking. 2017. *Computer Science and Information Systems* 14(1).

71. Yuqun Zeng, Liwei Wang, Hongfang Liu, Xusheng Liu, **Yanshan Wang***. Answering Diabetic Patients Questions Using Expert-ventted Online Resources: A Case Study. *IEEE International Conference on Bioinformation and biomedicine (BIBM)*, 2016. (* corresponding author)

72. **Yanshan Wang**, Majid Rastegar-Mojarad, Ravikumar Komandur-Elayavilli, Sijia Liu and Hongfang Liu. MayoNLPTeam at TREC 2016 Clinical Decision Support Track: An Ensemble Approach of Clinical Information Extraction and Retrieval, *Proceedings of the Text REtrieval Conference (TREC)*, 2016.

73. Dingcheng Li, Sijia Liu, Majid Mojarad Rastegar, **Yanshan Wang**, Xiaodi Li, Vipin Chaudhary, Terry Therneau, Hongfang Liu. A Topic-modeling Based Framework for Drug-drug Interaction Classification from Biomedical Text, *AMIA Annual Symposium*, 2016.

74. **Yanshan Wang**, Stephen Wu, Dingcheng Li, Saeed Mehrabi, and Hongfang Liu. "A Part-Of-Speech Term Weighting Scheme for Biomedical Information Retrieval." *Journal of Biomedical Informatics*, 2016.

75. **Yanshan Wang**, Stephen Wu, and Hongfang Liu. "MayoNLPTeam at the 2016 CLEF eHealth Information Retrieval Task.", *CLEF eHealth 2016*.

76. Sijia Liu, **Yanshan Wang**, Saeed Mehrabi, Dingcheng Li, Hongfang Liu. MayoBMI at ImageCLEF 2016 Handwritten Document Retrieval Task, *CLEF eHealth 2016*.

77. Vinod C. Kaggal, Ravikumar Komandur Elayavilli, Saeed Mehrabi, Joshua J. Pankratz, Sunghwan Sohn, **Yanshan Wang**, Dingcheng Li, Majid Mojarad Rastegar, Sean P. Murphy, Jason L. Ross, Rajeev Chaudhry, James D. Buntrock, Hongfang Liu. Towards Learning Healthcare System Knowledge Delivery at the Point of Care Empowered by Big Data and NLP. *Journal of Biomedical Informatics Insights*, 2016.

78. Naveed Afzal*, **Yanshan Wang***, Hongfang Liu. MayoNLP at SemEval-2016 Task 1: Semantic Textual Similarity based on Lexical Semantic Net and Deep Learning Semantic Model. Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016). Association for Computational Linguistics, 2016. [*co-first authors]

79. Saeed Mehrabi, **Yanshan Wang**, Donna Ihrke, Hongfang Liu. Exploring Gaps of Family History Documentation in EHR for Precision Medicine - A Case Study of Familial Hypercholesterolemia Ascertainment. *AMIA 2016 Joint Summits on Translational Science*, 2016.

80. **Yanshan Wang**, Saeed Mehrabi, Majid Rastegar Mojarad, Dingcheng Li, Hongfang Liu. Retrieval of Semantically Similar Healthcare Questions in Healthcare Forums. *the IEEE International Conference on Healthcare Informatics 2015 (ICHI 2015)*, 2015.

81. Dingcheng Li, Majid Rastegar-Mojarad, Ravikumar Komandur Elayavilli, **Yanshan Wang**, Saeed Mehrabi, Yue Yu, Sunghwan Sohn, Yanpeng Li, Naveed Afzal, Hongfang Liu. A Frequency-Filtering Strategy of Obtaining PHI-free Sentences

from Clinical Data Repository. *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics (ACM-BCB)*, 2015.

82. Dingcheng Li, Naveed Afzal, Majid Rastegar Mojarad, Ravikumar Komandur Elayavilli, Sijia Liu, **Yanshan Wang**, Feichen Shen, Hongfang Liu. Resolution of Chemical Disease Relations with Diverse Features and Rules. *Proceedings of the Fifth BioCreative Challenge Evaluation Workshop*, 2015.

83. **Yanshan Wang**, In-Chan Choi, Jae-Sung Lee. Indexing by Latent Dirichlet Allocation and Ensemble Model. *Journal of the American Society for Information Science and Technology (JASIS&T)*, 2014.

84. **Yanshan Wang**. Market index and stock price direction prediction using machine learning techniques: An empirical study on the KOSPI and HSI. 2014. *International Journal of Business Intelligence and Data Mining*, 9(2), 145-160.

85. **Yanshan Wang**. A novel soft keyboard for touch screen phones. 2013. *International Journal of Human Factors and Ergonomics*, 2(4), 246-261.

86. **Yanshan Wang**, In-Chan Choi. A Text Classification Method Based on Latent Topics. *Proceedings of the 1st International Conference on Operations Research and Enterprise Systems (ICORES 2012)*, 212-214, 2012.

PEER-REVIEWED ABSTRACTS

1. Bhavani Singh Agnikula Kshatriya, Nicolas A Nunez, Manuel Gardea-Resendez, Euijung Ryu, Brandon J Coombes, Sunyang Fu, Mark A Frye, Joanna M Biernacka, Yanshan Wang. Detecting Major Depressive Disorder from Clinical Notes using Neural Language Models with Distant Supervision. AMIA Annual Symposium, 2021.

2. Nicolas A Nunez, Joanna M Biernacka, Manuel Gardea-Resendez, Bhavani Singh Agnikula Kshatriya, Euijung Ryu, Sunyang Fu, Balwinder Singh, Brandon Coombes, Mark A Frye, Yanshan Wang. Natural Language Processing for Automatic Identification of Major Depressive Disorders in Free-Text Electronic Health Records. Society of Biological Psychiatry. 2021.

3. Luke A. Carlson, Jennifer St. Sauver, Sunyang Fu, Ahmad Tafti, Jungwei Fan, Molly M. Jeffery, Rozalina G. McCoy, Hongfang Liu, Yanshan Wang. Big Impact from Small Data: Unsupervised Machine Learning Approaches for Chronic Pain Patient Subgrouping. AMIA Annual Symposium, 2020.

4. Krishna B. Soundararajan, Sunyang Fu, Luke A. Carlson, Rebecca Smith, David S. Knopman, Hongfang Liu, Yanshan Wang. How Good is Artificial Intelligence at Automatically Answering Consumer Questions Related to Alzheimer's Disease? AMIA Informatics Summit, 2020.

5. Bradley A. White, Yanshan Wang, Michele L. Johnson, Ramona Lansing, Cadman L. Leggett, Liu Hongfang, and Prasad G. Iyer. Accuracy of natural language processing and bioinformatics approaches in identifying patient eligibility for Barrett's esophagus screening clinical trials. Gastroenterology 158, no. 6 (2020): S-306.

6. Yanshan Wang, Sunghwan Sohn. Trajectories of Functional Status of Cognitively-impaired Patients in EHRs. AMIA Annual Symposium, 2019.

7. Yanshan Wang, Andrew Wen, Sijia Liu, Jennifer St. Sauver, Adil E. Bharucha, Chunhua Weng, Hongfang Liu. Clinical Use of an Information Retrieval Framework for Cohort Discovery from Electronic Health Records. AMIA Annual Symposium, 2019.

8. Yiqing Zhao, Yanshan Wang, Paul R. Kingsbury, Michael Panzer, Richard Van Ert, Hongfang Liu. Summarization of Patient Education Material for New Media Platform. AMIA Annual Symposium, 2019.

9. Yiqing Zhao, Andrew Wen, Yanshan Wang, Feichen Shen, Suzette J. Bielinski, Hongfang Liu. Age as an Important Element in Analyzing Familial Pattern of Diseases. AMIA Informatics Summit, 2019.

10. Liwei Wang, Lei Luo, Yanshan Wang, Jason Wampfler A, Ping Yang, Hongfang Liu. Information Extraction for Data Population in Lung Cancer Clinical Research. AMIA Annual Symposium, 2018.

11. Yanshan Wang, Andrew Wen, Sijia Liu, Hongfang Liu. EMIRS: An Electronic Medical Information Retrieval System by Leveraging both Structured and Unstructured Electronic Health Records. AMIA Annual Symposium, 2018.

12. Naveed Afzal, Yanshan Wang, Feichen Shen, Liwei Wang, Hongfang Liu. A Resource for Clinical Semantic Textual Similarity. AMIA Annual Symposium, 2017.

13. Feichen Shen, Sijia Liu, Yanshan Wang, Liwei Wang, Naveed Afzal, Hongfang Liu. Accelerating Rare Disease Diagnosis with Collaborative Filtering. AMIA Annual Symposium, 2017.

14. Yanshan Wang, Naveed Afzal, Liwei Wang, Feichen Shen, Hongfang Liu. Semantic Representations of Medical Terms: A Comparison Study of Word Embeddings Trained on Clinical Notes and PubMed Articles. AMIA Annual Symposium, 2017.

15. Yuqun Zeng, Yanshan Wang, Feichen Shen, Sijia Liu, Liwei Wang, Majid Rastegar-Mojarad, Xusheng Liu, Hongfang Liu. Recommending education materials for diabetic questions using information retrieval approaches. AMIA Annual Symposium, 2017.

16. Dingcheng Li, Sijia Liu, Majid Rastegar-Mojarad, Yanshan Wang, Xiaodi Li, Vipin Chaudhary, Terry Therneau, Hongfang Liu. Drug-drug Interaction Detection with A Topic-modeling Based Framework Augmented with Distant-supervision. AMIA Annual Symposium, 2016.

17. Sunghwan Sohn, Yanshan Wang, Chung-Il Wi, Elizabeth A. Krusemark, Euijung Ryu, Mir H. Ali, Young J. Juhn, Hongfang Liu. Asthma Ascertainment NLP System Portability across Institutions. AMIA Annual Symposium, 2016.

18. Yuqun Zeng, Liwei Wang, Yanshan Wang, Dingcheng Li, Hongfang Liu. Answering patients' questions using expert-vetted online resources: A case study of diabetes. AMIA Annual Symposium, 2016.

19. Yanshan Wang, Stephen Wu, Dingcheng Li, Hongfang Liu. Influence of Part-of-Speech on the Clinical Information Retrieval, AMIA iHealth Clinical Informatics Conference, 2016.

20. Yanshan Wang, Stephen Wu, Dingcheng Li, Hongfang Liu. POS-MRF: A Part-Of-Speech Weighted Markov Random Field Model for Clinical Information Retrieval, AMIA 2016 Joint Summits on Translational Science, 2016.

21. Yanshan Wang, Dingcheng Li, Stephen Wu, Hongfang Liu. Improving Clinical Information Retrieval by Incorporating Part-Of-Speech Tagging. *Delivery Science Summit*, 2015.

SERVICE

Leadership:
- AMIA NLP Working Group Chair
- AMIA Awards Committee
- AMIA Nominating Committee
- AMIA Membership and Outreach Committee
- AMIA Membership and Outreach Committee

Organizer:
- HealthNLP 2018, 2019, 2020
- BioCreative/OHNLP Challenge 2018 Task 2: Clinical Semantic Textual Similarity
- National NLP Clinical Challenges (n2c2)/OHNLP Challenge 2019

Organization Committee:
- ICHI 2022

AMIA Awards Committee
2022, 2023

Editorial Member of the following journals/proceedings:
Frontiers in Artificial Intelligence; Biomedical Informatics Insights; Biomedical Informatics Insights Special Issue on Precision and Individualized Medicine; JMIR Medical Informatics Special Issue; MedInfo

Program Committee of of the following conferences:
AMIA; IEEE ICHI; BIOTECHNO; NLPCC; IHKDM; KDTBI; LREC; NAACL; EMNLP; COLING

Senior Program Committee of of the following conferences:
AMIA Informatics Summit

Student Paper Competition Committee:
AMIA 2017, 2018

Reviewer of the following journals:
Journal of Biomedical Informatics; Knowledge-Based Systems; Neurocomputing; Journal of Biomedical Engineering; International Journal of Image Mining; Journal of Multimedia Tools and Applications; Plos One; Applied Clinical Informatics; Journal of Medical Internet Research; SDRP Journal of Biomedical Engineering; Journal of Healthcare Informatics Research; Pharmaceutical Medicine; Nucleic Acids Research; Journal of Primary Care and Community Health; Journal of American Medical Informatics; Scientific Data; International Journal of Medical Informatics; Journal of Healthcare Informatics Research; IEEE Transactions on

Neural Networks and Learning Systems; IEEE Transactions on Knowledge and Data Engineering.

Reviewer of the following conferences:
COLING 2016-2020; EMNLP 2016-2020; ACM-BCB 2015-2020; BIBM 2015-2020; AMIA Joint Summits on Translational Science 2016-2020; AMIA Annual Symposium 2016-2020; International Workshop on Semantic Evaluation; SEPDA 2017; ICIBM 2017; HEALTHINFO 2017, 2018; BIOTECHNO 2017, 2018; SEPDA 2017; NAACL 2018-2020.

Session Chair of the following conferences:
HealthNLP 2018, 2019, 2020; CBMS 2020; ACM-BCB 2020.

Admission Committee:
Intelligent Systems Program, MSHI Program.

Talks
- Panelist. The American National Standards Institute (ANSI) Workshop on Data Standardization as Building Blocks for Artificial Intelligence (AI) in Healthcare. Nov. 30, 2021.
- Invited Talk. Computer Science Colloquium. Marquette University. Nov. 22, 2021.
- Keynote Speaker. The 6th International Workshop on Semantics-Powered Health Data Analytics (SEPDA 2021). Nov 5, 2021.
- Invited Talk. OHDSI NLP WG, Sep, 2021.
- Invited Talk. University of Pittsburgh, Feb, 2021.
- Invited Talk. Merck & Co., Jan, 2021.
- Invited Talk. United Health Group, Dec, 2020.
- Invited Talk. USF Moffitt Cancer Center, Dec, 2020.
- Keynote: Simple Introduction to Natural Language Processing and Its Clinical Applications in the Era of Artificial Intelligence, South Dakota State University Data Science Symposium, 2020.
- An Introduction to Natural Language Processing in Artificial Intelligence, Neuroscience Convergence 2019, 2019.
- CREATE: Cohort Retrieval Enhanced by Analysis of Text from Electronic Health Records, AMIA, 2018.
- BioCreative/OHNLP Challenge 2018, ACM-BCB, 2018.
- Towards A Distant Supervision Paradigm for Clinical Information Extraction: Creating Large Training Datasets for Machine Learning, South Dakota State University Data Science Symposium, 2018.
- Leveraging both Structured and Unstructured Data for Precision Information Retrieval, the Text REtrieval Conference (TREC), 2017.
- An Ensemble Approach of Clinical Information Extraction and Retrieval, the Text REtrieval Conference (TREC), 2016. [only 5 out of 24 teams were invited to give a presentation.]
- Systematic Analysis of Free-Text Family History in Electronic Health Record, AMIA Summits on Translational Science Proceedings
- MayoNLPTeam at TREC 2016 Clinical Decision Support Track: An Ensemble Approach of Clinical Information Extraction and Retrieval, the Text REtrieval Conference (TREC), Nov 17, 2016.
- Retrieval of Semantically Similar Healthcare Questions in Healthcare Forums, the IEEE International Conference on Healthcare Informatics, Oct 21, 2015.

- A Frequency-filtering Strategy of Obtaining PHI-free Sentences from Clinical Data Repository, the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics, Sep 10, 2015.
- Deep Structured Semantic Models, the Deep Learning Seminar, May 14, 2015.
- Document Ranking by Ensemble Models", Korea Business Intelligence and Data Mining Conference, 2013.

TEACHING

Courses

HI 2020 Practical Statistics & Programming Using Python and R, University of Pittsburgh   Fall 21-22

HI 2250 Foundations of Health Informatics, University of Pittsburgh   Fall 21-22

HI 2453 Data Analytics and Machine Learning in Health Science, University of Pittsburgh   Fall 21-22

Natural Language Processing & Health [Invited Speaker], Weill Cornell Medical College,   Spring 20-21

HINF 5610 (001) Foundations of Biomedical Natural Language Processing [Invited Speaker], University of Minnesota   Spring 19-20

Lectures/Tutorials

- RNN / LSTM Architectures and their applications in clinical note analytics, OSCT Annual Meeting: Deep Learning Foundation and Application with a Special Focus on Medical Informatics, May 6, 2019.
- Applications of Natural Language Processing in Clinical Research and Practice, the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), June 2, 2019.
- Methods and Applications of Natural Language Processing in Medicine, 2020 International Conference on Artificial Intelligence in Medicine (AIME), August 25, 2020.
- A Simple Introduction to Natural Language Processing and Its Clinical Applications in the Era of Artificial Intelligence. South Dakota State University Data Science Symposium, 2020.

FUNDED GRANTS

- NIH/Clinical and Translational Science Institute (CTSI), University of Pittsburgh 04/01/2022 – 03/31/2023
  Role: **Principal Investigator**
  Title: A3ST: AI-based Automated Fidelity Assessment for Strategy Training in Inpatient Rehabilitation
  Goal: To leverage advanced artificial intelligence (AI) techniques to automate the fidelity assessment approach, which have the potential to propel and translate to rehabilitation intervention practice and research forward in new directions previously untapped.
  Total Direct Costs: $25,000
- Pitt Momentum Fund, University of Pittsburgh   05/01/2022 – 04/31/2023
  Role: **Principal Investigator**
  Title: Improving Health Equity by Analyzing Social Determinants of Health from the Electronic Health Records
  Total Direct Costs: $25,000
- Year of Data and Society Funding, University of Pittsburgh   04/01/2022 – 03/31/2023
  Role: **Principal Investigator**

Title: Understanding Bias in Big Data and Artificial Intelligence for Health Care Through an Educational Health Informatics Hackathon
Total Direct Costs: $5,000

- AWS Diagnostic Development Initiative (DDI) Award, Amazon   10/01/2020 – 12/31/2022
  Role: **Principal Investigator**
  Title: Multimodal Machine Learning for Rapid Diagnosis
  Goal: This study tries to use CT images and de-identified electronic health records to develop machine learning or deep learning models for rapid disease diagnosis and related risk factors identification.
  AWS Credits: $20,000

- CHECE, Center for Health Equity and Community Engagement Research Award, Mayo Clinic   02/01/2020 – 01/31/2021
  Role: **Principal Investigator**
  Title: Developing Artificial Intelligence Models to Automatically Identify Social Determinants of Health Among Minority Populations from the Electronic Health Records and to Provide Implications for Health Equity
  Goal: This study has three goals; 1) attempt to automatically infer the presence of SDOH status of minority populations based on their EHRs; 2) develop and evaluate artificial intelligence (AI) models for inferring a patient's respective SBDH from EHR data; and 3) provide implications for health equity.
  Total Direct Costs: $50,000

- NIH/NIMH-R01MH121924   09/05/2019 – 05/31/2021
  Role: **Co-Investigator**
  Title: Leveraging EHR-linked biobanks for deep phenotyping, polygenic risk score modeling, and outcomes analysis in psychiatric disorders
  Goal: This proposal will apply big data techniques for development of polygenic risk scores and their association to clinical outcomes and social determinants using large-scale integrated phenotype-genotype data.

- NIH/NIKKD-R03DK128127   04/01/2021 – 05/31/2021
  Role: **Co-Investigator**
  Title: Digital Phenotyping of Nonalcoholic Fatty Liver Disease
  Goal: The objective of this study is to leverage data and analytics to 6 improve healthcare outcomes by early detection and risk stratification of NAFLD, before onset of liver-related 7 complications.

- NIH/NCATS-UL1TR02377   07/18/2017 – 06/30/2022
  Role: **Co-Investigator**
  Title: Mayo Clinic Center for Clinical and Translational Science (CCaTS)
  Goal: To facilitate the development of new therapies and their implementation in clinical practice; to train the next generation of clinical and translational physicians and scientists; to engage communities to fully participate in the translational research process; and to partner with regional and national networks to ultimately improve patient care and human health.

- NIH/NLM-R01LM011934   09/01/2014 – 07/31/2020
  Title: Semi-structured Information Retrieval in Clinical Text for Cohort Identification
  Goal: To make full use of clinical text in retrieving patients from the EMR. A layered language model for searching clinical text is introduced, addressing the

need for both fine-grained information and big-picture contextual information.

- NIH/NIH-UL1TR02377   08/01/2018 – 07/31/2019
  Role: **Co-Investigator**
  Title: Supplement Investigation of Chronic Pain Management Based on Electronic Health Records
  Goal: The long-term goal is to leverage the REP data and advanced informatics and analytics approaches to derive data-driven insights on chronic disease managements and build decision support tools for precision healthcare delivery.

- NIH/NIH-R01NS102233   06/01/2017 – 05/31/2021
  Role: **Co-Investigator**
  Title: Enabling Comparative Effectiveness Research in Silent Brain Infarction Through Natural Language Processing and Big Data
  Goal: The goal of this project is to improve the evidence base for the prevention of stroke in patients with silent brain infarct, i.e., a stroke on neuroimaging (head CT or MRI) but no clinical evidence of a stroke, using natural language processing that can accurately identify cases of silent brain infarction among a large population of adults.

CONSULTING SERVICES

Leverage AI techniques to improve the Clinical Practice in Mayo Clinic, the National #1 Hospital (Selected Projects)
- NLP to Enhance Patient Enrollment onto Cancer Clinical Trials
- AI to Predict Chemotherapy/Radiotherapy Outcomes using EHRs and Phenotyping Data
- AI for Smart Patient Scheduling
- NLP for Automatic Fracture Identification
- AI for Automatic Outside Materials Retrieval and Extraction
- Multimodal Machine Learning for Patient Satisfaction Assessment
- Chatbot for Automatic Healthcare Question-Answering
- Automatic Patient Education Materials Summarization