Yi***nin***g Hua

Boston, MA
yininghua@g.harvard.edu
+1 413-682-6681

Research Interest

AI for Healthcare, Mental Health Care, Epidemiology, Natural Language Processing

Education

Harvard T.H. Chan School of Public Health, Epidemiology (Ph.D.) Boston, MA
Advisors: Profs. Lori B. Chibnik, Marzyeh Ghassemi, and John Torous *Expected May 2028*

Harvard Medical School, Biomedical Informatics (MSc.) Boston, MA
Advisors: Profs. Li Zhou and David W. Bates *September 2021 - March 2023*

Harvard College, Undergraduate Visiting Program Cambridge, MA
Advisor: Dr. Chris Tanner *September 2020 - May 2021*

Smith College, Computer Science (B.A.) Northampton, MA
Advisor: Profs. Joseph O'Rourke and Jamie Macbeth *September 2017 - May 2021*

Research Experience

Division of Internal Medicine, Brigham and Women's Hospital Boston, MA
Research Assistant (I and II) *Oct. 2021 - Present*

- Supervisors: Li Zhou, PhD; David W. Bates, MD; Lauren V. Moran, MD; Dinah Foer, MD
- Study COVID-19 infoveillance using large-scale longitudinal Twitter data.
- Develop multimodal algorithms to identify patient cohorts from EHRs.

Boston Children's Hospital Boston, MA
Research Assistant II *Feb. 2023 - Aug. 2023*

- Supervisors: Scott HS. Wang, MD, PhD; Michael Lingzhi Li, PhD
- Used language models to extract multi-class labels/outcomes from urodynamic investigation charts.
- Examined efficiency of different active learning algorithms in multi-task and multi-class prediction.

Department of Dermatology, Massachusetts General Hospital Boston, MA
Research Trainee *Nov. 2021 - May 2022*

- Supervisors: Yevgeniy R. Semenov, MD, PhD
- Analyzed IBM Truven MarketScan data for health utilization studies.
- Examined image segmentation methods for melanoma stage II predictions.

Anesthesia, Beth Israel Deaconess Medical Center Boston, MA
Clinical Research Assistant II *Jun. - Sep. 2022*

- Supervisors: Maximilian S. Schaefer, MD; Haobo Ma, MD
- Developed algorithms for identifying delirium patient cohorts.
- Increased accuracy of patient identification from ICD codes (25% to 86%).

CELEHS, Harvard Medical School Boston, MA
Research Intern *May - Dec. 2021*

- Supervisors: Tianxi Cai, PhD
- Constructed cross-ontology hierarchical medical relations.
- Improved mapping accuracy by 30% over exact matching.

General Internal Medicine, Massachusetts General Hospital Boston, MA
Research Intern *Jun. - Nov. 2021*
- Supervisors: Jennifer S. Haas, MD
- Examined COVID-19's impact on the insurance status of breast cancer patients.

Computer Science, Smith College Northampton, MA
Research Assistant *May 2019 - Aug. 2020*
- Supervisors: Jamie C. Macbeth, PhD
- Developed an English Language Interpreter based on Conceptual Dependency theory.
- Compared performance of humans and pre-trained language models.

Selected Publications & Preprints (denotes equal contribution)*

Hua Y, Na H, Li Z, Liu F, Fang X, Clifton D, Torous J (2024). Applying and Evaluating Large Language Models in Mental Health Care: A Scoping Review of Human-Assessed Generative Tasks. *doi: 10.48550/arXiv.2408.11288*

Hua Y, Xia W, Bates D W, Hartstein L, Kim H T, Li M L, Nelson B W, Stromeyer C IV, King D, Suh J, Zhou L, Torous J (2024). Standardizing and Scaffolding Healthcare AI-Chatbot Evaluation. *doi: 10.1101/2024.07.21.24310774*

Liu A, Zhou H, **Hua Y**, Rohanian O, Clifton L, Clifton A (2024). Large Language Models in Healthcare: A Comprehensive Benchmark. *arXiv preprint arXiv:2405.00716*

Hua Y, Blackley S, Shinn A, Skinner J, Moran L, Zhou L (2024). Identifying Psychosis Episodes in Psychiatric Admission Notes via Rule-based Methods, Machine Learning, and Pre-Trained Language Models. *doi: 10.21203/rs.3.rs-4126574/v1*

Hua Y, Mukkamala A, Estrada C, Li M, Wang H (2024). High-performing Multi-task Model of Urinary Tract Dilation (UTD) Classification for Neonatal Ultrasound Reports Through Natural Language Processing. *doi: 10.1101/2024.01.23.24301680*

Hua Y, Lin S, Li M, Zhang Y, Foer D, Zhou P, Zhou L, Yang J. Streamlining Social Media Information Retrieval for COVID-19 Research with Deep Learning. J Am Med Inform Assoc. 2024 Jun 20;31(7):1569-1577

Hua Y, Liu F, Yang K, Li Z, Na H, Sheu Y, Zhou P, Moran L, Ananiadou S, Beam A, Torous J (2024). Large Language Models in Mental Health Care: a Scoping Review. *arXiv: 2401.02984*

Zhou H, Gu B, Zou X, Li Y, Chen S S, Zhou P, Liu J, **Hua Y**, Mao C, Wu X, Li Z, Liu F (2023). A Survey of Large Language Models in Medicine: Progress, Application, and Challenge. *arXiv: 2311.05112*

Liu J*, Zhou P*, **Hua Y***, et al. (2023). Benchmarking Large Language Models on CMExam–A Comprehensive Chinese Medical Exam Dataset. *Advances in Neural Information Processing Systems (NeurIPS)*, 36

Zeng Q, Garay L, Zhou P, Chong D, **Hua Y**, Wu J, Pan Y, Zhou H, Voigt R, Yang J. GreenPLM: Cross-lingual transfer of monolingual pre-trained language models at almost no cost. *The 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, 2023

Hua Y, Wang L, Nguyen V, Rieu-Werden M, McDowell A, Bates D W, Zhou L. A Deep Learning Approach for Transgender and Gender Diverse Patient Identification in Electronic Health Records. *J Biomed Inform.* 2023 Nov;147:104507.

Wu J*, Wang L*, **Hua Y**, Li M, Zhou L, Bates D W, Yang J. Trend and Co-occurrence Network of COVID-19 Symptoms From Large-Scale Social Media Data: Infoveillance Study. *J Med Internet Res.* 2023 Mar 14;25:e45419.

Wu J, Wu X, **Hua Y**, Lin S, Zheng Y, Yang J. Exploring Social Media for Early Detection of Depression in COVID-19 Patients. *Proceedings of the ACM Web Conference (WWW)*, 2023

Liu J*, Wang Z*, Ye Q*, Chong D*, Zhou P*, **Hua Y***. Qilin-Med-VL: Towards Chinese Large Vision-Language Model for General Healthcare. *arXiv:2310.17956*

Zhou P, Wang Z, Chong D, Guo Z, **Hua Y**, Su Z, Teng Z, Wu J, Yang J. METS-CoV: A Dataset of Medical Entity and Targeted Sentiment on COVID-19 Related Tweets. *Advances in Neural Information Processing Systems (NeurIPS)*, 35, 21916-21932

Li M*, **Hua Y***, Liao Y, Zhou L, Li X, Wang L, Yang J. Tracking the Impact of COVID-19 and Lockdown Policies on Public Mental Health Using Social Media: Infoveillance Study. *J Med Internet Res.* 2022 Oct 13;24(10):e39676.

Hua Y, Jiang H, Lin S, Yang J, Plasek J M, Bates D W, Zhou L. Using Twitter data to understand public perceptions of approved versus off-label use for COVID-19-related medications. *J Am Med Inform Assoc.* 2022 Sep 12;29(10):1668-1678.

Wan G, Nguyen N, Liu F, DeSimone M S, Leung B W, Rajeh A, Collier M R, . . . , **Hua Y**, et al. Prediction of early-stage melanoma recurrence using clinical and histopathologic features. *NPJ Precis Oncol.* 2022 Oct 31;6(1):79.

Conference Talks (as the presenting author)

AMIA Informatics Summit, Boston, MA. *Identifying Psychosis Episodes in Admission Notes: A Comparative Study of Rule-based Methods, Machine Learning Algorithms, and Pre-Trained Language Models*. March 2024.

AMIA Informatics Summit, Boston, MA. *Wise Instance Selection Algorithms Help Reduce Annotation Work in Multi-Task Multi-Class Urinary Tract Dilation Prediction*. March 2024.

AMIA Informatics Summit, Boston, MA. *One Model Fits All? High-performing Multi-task Model of Urinary Tract Dilation (UTD) Classification Using NLP for Neonatal Ultrasound Reports*. March 2024.

Conference on Neural Information Processing Systems (NeurIPS), New Orleans, LA. *Benchmarking Large Language Models on CMExam–A Comprehensive Chinese Medical Exam Dataset*. December 2023.

AMIA Annual Symposium, New Orleans, LA. *Streamlining Social Media Information Retrieval for Public Health Research with Deep Learning*. November 2023.

IEEE International Conference on Healthcare Informatics (ICHI) , Houston, TX. *Streamlining Social Media Information Retrieval for Public Health Research with Deep Learning*. June 2023.

IEEE International Conference on Healthcare Informatics, Houston, TX. *A Deep Learning-driven Approach for Transgender and Gender Diverse Patient Identification in EHRs*. June 2023.

American Academy of Neurology Annual Meeting, Boston, MA. *Contemporary treatment and outcomes of myasthenia gravis in the United States*. April 2023.

American Medical Informatics Association (AMIA) Annual Symposium, Washington, D.C. *Identification of transgender and gender diverse individuals in electronic health records*. November 2022.

AMIA Annual Symposium, Washington, D.C. *Using Twitter to Understand Public Perceptions of COVID-19 Drugs*. November 2022.

Asia-Pacific Babylab Constellation Conference, Hong Kong. *Quantifying the bilingual (dis) advantage in vocabulary acquisition*. December 2021.

Teaching Services

PHS 2000A: Quant. Res. Methods Pop. Health Sci. Harvard School of Public Health
Teaching Fellow *Fall 2024*

BMI 707/EPI-290: Deep Learning for Biomedical Data Harvard Medical School
Teaching Fellow *Spring 2023*

S-043/Stat-151: Multilevel and Longitudinal Models Harvard School of Education
Teaching Fellow *Summer 2021*

Machine Learning for Self-Driving Cars Harvard School of Public Health
Teaching Assistant *Summer 2020*

PHY 215: Light, Relativity, and Quantum Physics Smith College
Teaching Assistant *Spring 2020*

PHY 210: Math Methods of Physical Sciences and Engineering Smith College
Teaching Assistant *Fall 2019*

Awards & Accolades

The 11th IEEE ICHI 2023
Best poster

The 36th Conference on NeurIPS 2022
Spotlight long paper

The AMIA Annual Symposium 2022
Distinguished poster

The Sigma Xi honor society 2021
Nominated outstanding student researcher

Fundings & Fellowships

Harvard T.H. Chan School of Public Health 2024
The Rose Traveling Fellowship

Harvard T.H. Chan School of Public Health 2023
The Brian and Heidi MacMahon Epidemiology Educational Fund

The 37th Conference on NeurIPS 2023
Travel Fund

The Grace Hopper Celebration of Women in Computing 2020
Student scholarship

Smith College 2020
Project Coach fellowship

Smith College 2019
Praxis fellowship

Review Services

npj Digital Medicine 2024
IF: 15.357

JMIR Public Health and Surveillance 2024
IF: 14.56

JMIR Mental Health 2024
IF: 5.2

Journal of Medical Internet Research (JMIR) 2024
IF: 7.6

The Journal of American Medical Informatics Association (JAMIA) 2023, 2024
IF: 7.942

Machine Learning for Health (ML4H) Symposium 2023

The AMIA Informatics Summit 2023

The Conference on Neural Information Processing Systems (NeurIPS) 2023

The International Journal of Medical Informatics (IJMI) 2022, 2024

The 29th International Conference on Computational Linguistics (COLING) 2022

The AMIA Annual Symposium 2022, 2023