OMB No. 0925-0001 and 0925-0002 (Rev. 10/2021 Approved Through 09/30/2024)

**BIOGRAPHICAL SKETCH**

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

NAME: YE, YE

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

POSITION TITLE: Postdoc Associate, Department of Biomedical Informatics, University of Pittsburgh

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

| INSTITUTION AND LOCATION | DEGREE*(if applicable)* | Completion DateMM/YYYY | FIELD OF STUDY |
| --- | --- | --- | --- |
| Peking University, Beijing, China | BMED | 07/2006 | Preventive Medicine |
| Peking University, Beijing, China | MS | 01/2009 | Epidemiology and Health Statistics |
| Emory University, Atlanta, GA | MSPH | 05/2011 | Public Health Informatics |
| University of Pittsburgh, Pittsburgh, PA | PhD | 12/2018 | Intelligent Systems |
| University of Pittsburgh, Pittsburgh, PA | Postdoc | 04/2022 | Biomedical Informatics |

**A. Personal Statement**

My research mainly focuses on developing artificial intelligence technologies for biomedicine. I have developed and applied machine learning and natural language processing algorithms and systems for automated infectious disease surveillance using hospital electronic medical records and nationwide over-the-counter products, patient readmission risk profiling, and tweets-based adverse drug reaction monitoring.

During my PhD study, I worked as a graduate student researcher for **R01 LM011370 Probabilistic Disease Surveillance**. I used natural language processing to retrieve clinical features from free-text medical reports, developed machine learning techniques to generate multiple case detection systems (CDS), and tested different configurations, including chief complaint based CDS, dynamic disease prior, and dynamic clinical data. Moreover, I used the developed CDS systems to compute likelihoods of having diseases, which enabled the outbreak detection and characterization system component to conduct population-level probabilistic outbreak detection and characterization.

My most recent work involves studying the transfer of computable biomedical knowledge from one setting to another. I successfully demonstrated high inherent transferability (without transfer learning) of an Allegheny County (PA)-developed case detection system to a Salt Lake County (UT) system and moderate transferability of a Salt Lake County-developed case detection system to an Allegheny County system. This success indicates the feasibility of sharing case detection models from one region to another and at the same time suggests the need for techniques to improve transferability when performance drops; this became the focus of my doctoral dissertation. My dissertation developed data-based and model-based transfer learning algorithms, both of which were shown to work well on both synthetic scenarios and influenza case detection scenarios. This dissertation is on the finalist list (of highest five scored dissertations) for the 2019 American Medical Informatics Association Doctoral Dissertation Competition.

In May 2020, my grant, “**Transfer learning to improve the re-usability of computable biomedical knowledge**,” was awarded funding by the NIH National Library of Medicine (**K99LM013383**). In May 2022, I became an assistant professor and continuous the research grant as P.I. (R00 stage), with which I am conducting transfer learning techniques-related research for infectious diseases. Meanwhile, I am also collaborating with other researchers on **R01LM013509 Automated Surveillance of Overlapping Outbreaks and New Outbreak Diseases** and **CDC U24 OH009077 National Mesothelioma Virtual Bank.**

**B. Positions, Scientific Appointments and Honors**

**Positions and Employment**

|  |  |
| --- | --- |
| May 2022 - present | Assistant Professor, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA |
| 2019 - Apr. 2022 | Postdoc Associate, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA |
| 2018 | PhD Candidate Full-time Dissertation Study, Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA |
| 2016 | Teaching Assistant, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA |
| 2011 - 2017 | Graduate Student Researcher, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA |
| 2010 - 2011 | Research Assistant, Department of Biostatistics & Bioinformatics, Emory University, Atlanta, GA  |
| 2010 | Intern, Centers for Disease Control and Prevention, Atlanta, GA |
| 2006 - 2008 | Research Assistant, Department of Epidemiology & Biostatistics, Peking University, Beijing, China |

**Other Experience**

|  |  |
| --- | --- |
| 2017 - Present | Reviewer. *International Journal of Medical Informatics* |
| 2017 - Present | Reviewer. *Journal of Public Health Management and Practice* |
| 2017 - Present | Reviewer. *Artificial Intelligence in Medicine* |
| 2017 - Present | Reviewer. *Computers in Biology and Medicine* |
| 2018 - Present | Reviewer. *The BMJ* |
| 2018 - Present | Reviewer. *Online Journal of Public Health Informatics* |
| 2020 - Present | Reviewer. *Journal of the American Medical Informatics Association* |
| 2020 - Present | Reviewer*. Journal of Biomedical Informatics* |
| 2017 | Reviewer. *Journal of Innovation in Health Informatics* |
| 2014, 2015, 2017 | Reviewer. American Medical Informatics Association Annual Symposium |

**Professional Memberships**

|  |  |
| --- | --- |
| 2024 - 2027 | Journal Editorial Board. Journal of Biomedical Informatics |
| 2022 | Program Committee. The Eighth International Workshop on Data Management and Analytics for Medicine and Healthcare |
| 2018 - 2019 | Member. Association for the Advancement of Artificial Intelligence |
| 2018 | Journal Editorial Board. *Computers in Biology and Medicine* |
| 2017 - Present | Journal Editorial Board. *Artificial Intelligence in Medicine* |
| 2016 - Present | Member. Observational Health Data Sciences and Informatics Program |
| 2010 - Present | Member. American Medical Informatics Association |

**Honors**

|  |  |
| --- | --- |
| 2019 | American Medical Informatics Association Doctoral Dissertation Award Finalist |
| 2017 | Best Student Paper 2nd Place, Department of Biomedical Informatics, University of Pittsburgh |
| 2016 | Innovator Award, University of Pittsburgh |
| 2015 | Distinguished Poster Award (3rd author), American Medical Informatics Association Annual Symposium  |
| 2014 | Distinguished Poster Award (2nd author), American Medical Informatics Association Annual Symposium  |
| 2014 | Best Statistical Science Theoretical Paper (3rd author), Centers for Disease Control and Prevention |
| 2014 | Best Student Paper 2nd Place, Department of Biomedical Informatics, University of Pittsburgh |
| 2013 | Finalist in Michael G. Wells Student Health Care Entrepreneurship Competition, University of Pittsburgh |
| 2011 | Delta Omega Honor Society in Public Health |

**C. Contributions to Science**

**C1. Transfer Learning**

As more and more machine-learned models are generated, the biggest challenge is that a model developed in one healthcare system (denoted as source) may be expected to underperform in another healthcare system (target). I developed an innovative transfer learning framework to enable sharing of both data and models between source and target. My research enables knowledge-sharing under heterogeneous scenarios and provides an approach for understanding transfer learning performance between source and target in terms of differences of features and their distributions and sample sizes. The model-transfer algorithm can be viewed as a new Bayesian network learning algorithm that has a flexible representation of prior knowledge. I am now continuing research on Bayesian and deep learning approaches to improve the re-usability of models. Re-using models is expected to benefit the public’s health by: (1) improving case detection during epidemics by enabling re-use of automatic case detectors developed in the earliest affected regions with other regions, and, more generally, (2) increasing the impact of NIH’s investment in machine learning by enabling machine-learned models to be used in more institutions and locations.

1. Ji Y, Gao Y, Bao R, Li Q, Liu D, Sun Y, **Ye Y\***. Prediction of COVID-19 Patients' Emergency Room Revisit using Multi-Source Transfer Learning. https://arxiv.org/abs/2306.17257. 2023 Jun 29. (Accepted and presented in IEEE ICHI) (Corresponding author)
2. **Ye Y**, Wagner MM, Cooper GF, Ferraro JP, Su H, Gesteland PH, Haug PJ, Millett NE, Aronis JM, Nowalk AJ, Ruiz VM, López Pineda A, Shi L, Van Bree R, Ginter T, Tsui F. A study of the transferability of influenza case detection systems between two large healthcare systems. PLoS One. 2017;12(4):e0174970. doi: 10.1371/journal.pone.0174970. eCollection 2017. PubMed PMID: 28380048; PubMed Central PMCID: PMC5381795.
3. **Ye Y**. Transfer Learning for Bayesian Case Detection Systems. Doctoral dissertation, University of Pittsburgh. 2018. \*American Medical Informatics Association Doctoral Dissertation Award Finalist (Top 5) in 2019.
4. **Ye Y**, Wagner MM, Cooper GF, Tsui FC, Weiss J, Gesteland PH, Jeffrey Ferraro JP, Haug PJ. Bayesian network transfer learning to improve re-usability of computable biomedical knowledge for public health. Poster abstract accepted by Mobilizing Computable Biomedical Knowledge 2nd Annual Meeting in National Institutes of Health - Bethesda, MD, July 18-19, 2019.
5. Ferraro JP, **Ye Y**, Gesteland PH, Haug PJ, Tsui FR, Cooper GF, Van Bree R, Ginter T, Nowalk AJ, Wagner M. The effects of natural language processing on cross-institutional portability of influenza case detection for disease surveillance. Appl Clin Inform. 2017 May 31; 8(2):560-580. doi: 10.4338/ACI-2016-12-RA-0211. PubMed PMID: 28561130; PubMed Central PMCID: PMC6241736.

**C2. Public Health Surveillance**

Traditionally, public health surveillance relies mainly on sentinel physician reporting and laboratory reporting, which usually result in reporting delays and issues due to underreporting and undertesting. I have developed a near real-time, active surveillance approach that uses machine-learned case detection systems to automatically capture cases from electronic medical records. A case detection system consists of a natural language processing parser (NLP) and a Bayesian network classifier. It uses NLP to infer the presence or absence of clinical findings from narrative notes. With these findings, a Bayesian network classifier infers each patient’s diagnosis probabilities, as well as the likelihood of patient clinical evidence, to support outbreak detection and forecast at the population-level. My case detection systems enhance the communications between clinicians and public health officials, automatically inferring patients’ diagnoses from free-text clinical notes for population monitoring and automatically providing population prevalence information to support clinical decisions on differential diagnoses. My research shows the impact of accurate natural language processing and feature selection on classification performance and demonstrates the advantages of automated predictive modeling versus experts’ simple judgement with regards to making correlations between NLP-parsed clinical findings and disease status. Regarding model parameterization, my experiments also show that using a changing prevalence of disease could increase the discriminative ability of an influenza detection model compared to using a constant prevalence. Moreover, I demonstrated high performance for influenza detection in a five-year retrospective study in both Allegheny County and Salt Lake County.

1. **Ye Y**, Tsui FR, Wagner M, Espino JU, Li Q. Influenza detection from emergency department reports using natural language processing and Bayesian network classifiers. **J Am Med Inform Assoc**. 2014 Sep-Oct; 21(5):815-23. doi: 10.1136/amiajnl-2013-001934. Epub 2014 Jan 9. PubMed PMID: 24406261; PubMed Central PMCID: PMC4147621.
2. López Pineda A, **Ye Y**, Visweswaran S, Cooper GF, Wagner MM, Tsui FR. Comparison of machine learning classifiers for influenza detection from emergency department free-text reports. **J Biomed Inform**. 2015 Dec; 58:60-69. doi: 10.1016/j.jbi.2015.08.019. Epub 2015 Sep 16. PubMed PMID: 26385375; PubMed Central PMCID: PMC4684714.
3. Tsui F, **Ye Y**, Ruiz V, Cooper GF, Wagner MM. Automated influenza case detection for public health surveillance and clinical diagnosis using dynamic influenza prevalence method. **J Public Health** (Oxf). 2018 Dec 1; 40(4):878-885. doi: 10.1093/pubmed/fdx141. PubMed PMID: 29059331; PubMed Central PMCID: PMC6676953.
4. Aronis JM, Millett NE, Wagner MM, Tsui F, **Ye Y**, Ferraro JP, Haug PJ, Gesteland PH, Cooper GF. A Bayesian system to detect and characterize overlapping outbreaks. **J Biomed Inform**. 2017 Sep; 73:171-181. doi: 10.1016/j.jbi.2017.08.003. Epub 2017 Aug 7. PubMed PMID: 28797710; PubMed Central PMCID: PMC5604259.

**C3. Automated Readmission Risk Profiling**

Unplanned hospital readmissions are placing an unbearable financial burden on the country. We developed readmission risk assessment models for heart failure for adults and children by analyzing unstructured data (e.g., discharge reports) and structured data (e.g., laboratory orders and results, drug orders, healthcare utility information) stored in hospital electronic medical record systems. We developed a domain ontology-based feature selection algorithm that performs better than feature selection algorithms without ontology knowledge. Our work on automated readmission risk profiling received Distinguished Poster Awards at the 2014 and 2015 American Medical Informatics Annual Symposium. In addition, three main personnel, including myself, received an innovator award from the University of Pittsburgh for this work. Furthermore, our pediatric readmission prediction model has been integrated into the Cerner® Millennium electronic medical record system at the Children’s Hospital of Pittsburgh of UPMC. Both retrospective evaluation and prospective evaluation indicate that the system has the potential to identify patients with elevated risk of readmission, thereby allowing more informed and more successful discharge planning and intervention.

1. Lu S, **Ye Y**, Tsui R, Liu X, Hwa R. Feature selection for 30-day heart failure readmission prediction using clinical drug data. *NIPS Workshop on Machine Learning for Clinical Data Analysis and Healthcare*, Harrahs and Harveys, Lake Tahoe, 2013.
2. Lu S, **Ye Y**, Tsui R, Su H, Rexit R, Wesaratchakit S, Liu X, Hwa R. Domain ontology-based feature reduction for high dimensional drug data and its application to 30-day heart failure readmission prediction. *In Collaborative Computing: Networking, Applications and Worksharing (Collaboratecom), 9th International Conference* pp. 478-484. IEEE. 2013, Oct 20
3. Draper AJ, **Ye Y**, Ruiz VM, Patterson C, Urbach A, Palmer F, Wang S, Somboonna M, and Tsui F. Using laboratory data for prediction of 30-day hospital readmission of pediatric seizure patients. *2014 AMIA* paper abstract, Washington, DC. \*Distinguished Poster Award
4. Tsui F, Ruiz V, Barda A, **Ye Y**, Butler G, Suresh S, Urbach A, Retrospective and prospective evaluations of the system for hospital adaptive readmission prediction and management (SHARP) for all-cause 30-Day pediatric readmission prediction. *2017 AMIA* paper abstract.

**C4. Challenges, Opportunities of Artificial Intelligence in Biomedicine**

Growing numbers of artificial intelligence applications are being developed and applied to biomedicine. These technologies introduce risks and benefits that must be assessed and managed. In the first article, we discussed how long-standing principles of medical and scientific ethics can be applied to artificial intelligence. In the second article, we focus on the role of AI in clinical and translational research (CTR), including preclinical research, clinical research, clinical implementation, and public (or population) health. For each CTR phase, we addressed challenges, successes, failures, and opportunities for AI. We present three complementary perspectives: (1) scoping literature review, (2) survey, and (3) analysis of federally funded projects. In this third article, We reviewed the existing sustainability models for open-source software (OSS) and describes 10 OSS use cases, including 3D Slicer, Bioconductor, Cytoscape, Globus, i2b2 (Informatics for Integrating Biology and the Bedside) and tranSMART, Insight Toolkit, Linux, Observational Health Data Sciences and Informatics tools, R, and REDCap (Research Electronic Data Capture), in 10 sustainability aspects: governance, documentation, code quality, support, ecosystem collaboration, security, legal, finance, marketing, and dependency hygiene.

1. Jackson BR, **Ye Y**, Crawford JM, Becich MJ, Roy S, Botkin JR, de Baca ME, Pantanowitz L. The ethics of artificial intelligence in pathology and laboratory medicine: principles and practice. Academic Pathology. 2021 Feb 15;8:2374289521990784.
2. Bernstam EV, Shireman PK, Meric‐Bernstam F, N Zozus M, Jiang X, Brimhall BB, Windham AK, Schmidt S, Visweswaran S, **Ye Y**, Goodrum H. Artificial intelligence in clinical and translational science: Successes, challenges and opportunities. Clinical and translational science. 2021 Oct 27.
3. **Ye Y**, Barapatre S, Davis MK, Elliston KO, Davatzikos C, Fedorov A, Fillion-Robin JC, Foster I, Gilbertson JR, Lasso A, Miller JV. Open-source Software Sustainability Models: Initial White Paper From the Informatics Technology for Cancer Research Sustainability and Industry Partnership Working Group. Journal of medical Internet research. 2021 Dec 2;23(12):e20028.

**C5. Evaluation of Information Systems**

The development of a medical information system must include comprehensive evaluation of its successes and deficiencies when applied in real-world scenarios. During my internship at the U.S. CDC, I designed an evaluation plan to study the effectiveness of an EMR alerting service. This plan showed how to take into account the objectives of major stakeholders, select an evaluation framework, identify evaluation elements, and clarify measurement methods. In the plan, I recommend the evaluation of causal relationships among input, process, and outcomes, as well as correlations between objective and subjective indicators. These causal relationships can indicate the validity of many measurements. This evaluation could be easily adapted for evaluation of other information systems.

1. **Ye Y**. An evaluation plan for a pilot of CDC EMR alerting service prototype. Master’s thesis, Emory University. 2011.