XINSONG DU

Personal Website: https://bit.ly/3yVgkay
Google Scholar: https://bit.ly/3ka7nnP
GitHub: https://github.com/XinsongDu
Email: xinsongdu@ufl.edu

2901 SW 13TH ST APT 313
Gainesville, FL 32608

I am a Ph.D. candidate and Graduate Research Assistant in Department of Health Outcomes and Biomedical Informatics at University of Florida (UF), as well as a UF Informatics Institute Fellow. My research interests include biomedical informatics, computational metabolomics, machine learning, research software development and research reproducibility improvement. My technical skills include Python, Shell, R, Nextflow/Groovy, etc. My career goal is to become an externally funded investigator focusing on biomedical informatics and health outcomes research.

EDUCATION

Ph.D. University of Florida, Biomedical Informatics — Aug 2017 – Present

Supervisory Committee:

- Dr. Dominick J. Lemas (Chair)
- Dr. William R. Hogan
- Dr. Timothy J. Garrett
- Dr. Mathias Brochhausen
- Dr. Mei Liu

Training:

- Mass-spectrometry
- Metabolomics
- Research software development
- Natural language processing
- Research reproducibility
- Biomedical sample handling
- Programming: Nextflow/Groovy

M.S. University of Florida, Computer Engineering — Aug 2015 – May 2017

Training:

- Data analysis for electronic health records
- Machine Learning
- High-performance computing
- Computer vision
- Programming: Python/R/Shell
- Computer architecture
- Computer communications

B.S. Shandong University, Electrical Engineering — Sep 2011 – Jul 2015

Training:

- Programming: C/C++/Assembly Language/MATLAB/Action Script
- Mathematics: Calculus, statistics, linear algebra, field theory, signal processing
- Single-chip micro computer
- Physics: electronic magnetic field
- Power electronics

RESEARCH FUNDING

UF Informatics Institute Graduate Student Fellowship 04/2021-04/2023

Support dissertation research *Enabling Reproducible Untargeted Metabolomics Research: Next Generation Untargeted Metabolomics Data Analysis Workflow* ($45,000)
Research funding given to selected graduate students nominated by their department who are working on informatics-related area at UF (awards about 5 graduate students in total each year).
University of Florida.

Graduate Research Assistantship 2017-Present
Department of Health Outcomes and Biomedical Informatics,
College of Medicine,
University of Florida.

HONORS AND AWARDS

UF Three Minute Thesis (3MT) Competition Finalist 2022
Ranked top-10 in the competition.

Alec Courtelis Award (Nomination) 2022
Award for international students who contributed significantly to the community and have outstanding academic achievement.
Each college can nominate a maximum of two students every year.

Certificate of Outstanding Merit 2022
Certificate issued to selected international students,
UF College of Medicine and International Center

Study Florida & FAIE 2022 High Education Scholarship (Nomination) 2021
Scholarship that awards international students studying in Florida.
A maximum of three international students can be nominated by each university every year.

Certificate of Outstanding Merit 2021
Certificate issued to selected international students,
UF College of Medicine and International Center

Achievement Award Scholarship for New Engineering Graduate Students 2015
Award given to selected incoming graduate student at College of Engineering,
University of Florida.

RESEARCH EXPERIENCE

Department of Health Outcomes & Biomedical Informatics
University of Florida, Gainesville, FL 08/2017 - Present
Graduate Research Assistant
Mentor: Dr. Dominick J. Lemas

Department of Electrical & Computer Engineering
Department of Surgery
University of Florida, Gainesville, FL 06/2016 - 05/2017
Research Assistant
Mentors: Dr. Azra Bihoric and Dr. Xiaolin Andy Li

Department of Electrical Engineering
Shandong University, Jinan, China 09/2014 - 04/2015
Undergraduate Researcher
Mentor: Dr. Guangzhu Wang

PUBLICATIONS

Journal Publications

Du. X.; Dastmalchi. F.; Ye. H.; Garrett. T.J.; Dillar. M.A.; Liu. M.; Hogan. W.R.; Brochhausen. M.; Lemas. D.J.. Evaluating LC-HRMS Metabolomics Data Processing Software using FAIR Principles for Research Software. *Metabolomics* (Editor-Selected Cover Article). 2023. PMID: 36745241

Du, X.; Aristizabal-Henal, J.J.; Garrett, T.J.; Brochhausen, M.; Hogan, W.R.; Lemas, D.J. A checklist for reproducible computational analysis in clinical metabolomics research. *Metabolites*. 2021. PMID: 35050209

Lemas, D.J.; Wright, L.; Flood-Grady, E.; Francois, M.; Chen, L.Y.; Hentschel, A.; **Du, X.**; Hsiao, C.J.; Chen, H.; Neu, J.; Theis, R.P.; Shenkman, E.; Krieger, J. Perspectives of pregnant and breastfeeding women on longitudinal clinical studies that require non-invasive biospecimen collection – a qualitative study. *BMC Pregnancy and Childbirth*. 2021. PMID: 33472584

Hentschel, A.; Hsiao, C.J.; Chen, L.Y.; Wright, L.; Shaw, J.; **Du, X.**; Flood-Grady, E.; Harle, C.A.; Reeder, C.F.; Francois, M.; Louis-Jacques, A.; Shenkman, E.; Krieger, J.; Lemas, D.J. (2021) Perspectives of pregnant and breastfeeding persons on participating in longitudinal mother-baby studies involving electronic health records: a qualitative study. *JMIR Pediatrics and Parenting.* 2021. PMID: 33472584

Lemas, D.J.; Loop, L.S.; Duong, M.; Schleffer, A.; Collins, C.; Bowden, J.A.; **Du, X.**; Patel, K.; Ciesielski, A.L.; Ridge, Z.; Wagner, J.; Subedi, B.; Delcher, C. Estimating drug consumption

during a college sporting event from wastewater using liquid chromatography mass spectrometry. *Science of The Total Environment*. 2021. PMID: 33385644

Lure, A.C.; **Du, X.**; Black E.W.; Irons, R.; Lemas, D.J.; Taylor, J.A.; Lavilla, O.; de la Cruz, D.; Neu, J. Using machine learning analysis to assist in differentiating between necrotizing enterocolitis and spontaneous intestinal perforation: a novel predictive analytics tool. *Journal of Pediatric Surgery*. 2020. PMID: 33342603

Du, X.; Min, J.; Shah C.P.; Boshnoi, R.; Hogan W.R.; Lemas, D.J. Predicting in-hospital mortality of patients with febrile neutropenia using machine learning models. *International Journal of Medical Informatics*. 2020. PMID: 32325370

Bian, J.; Zhao, Y.; Salloum, R.G.; Guo, T.; Wang, M.; Prosperi, M.; Zhang, H.; **Du, X.**; Ramirez-Diaz, L.J.; He, Z.; Sun, Y. (2017) Using social media data to understand the impact of promotional information on laypeople's discussions: a case study of lynch syndrome. *Journal of Medical Internet Research*. 2017. PMID: 29237586

Du, C.; **Du, X.** (2016) Cache optimization by fully-replacement policy. *American Journal of Embedded Systems and Applications.* 2016.

Conference Papers

(Peer-Reviewed)

Du, X.; Bian, J.; Prosperi, M. An operational deep learning pipeline for classifying life events from individual tweets. *5th International Conference on Information Management and Big Data*, Sep. 03-05, 2018. **(Oral)**

(Abstract-Reviewed)

Du. X.; Dastmalchi. F.; Diller. M.A.; Brochhausen. M.; Garrett. T.J.; Hogan. W.R.; Lemas. D.J. (2023). An Automated Workflow Composition System for LC-MS Metabolomics Research. Annual Conference of American Society of Mass Spectrometry. Jun 04-08, 2023. Houston, TX. **(Oral)**

Wane, I.; Elias, E.; Xu, K.; **Du, X.**; Bisesi, J.H.; Young, H.W.; Brown, K., Bowden, J.A., Delcher, C., Lemas, D.J. Utilization of wastewater-bsed epidemiology to identify drug consumption and predict health outcomes using electronic health records. Annual Biomedical Research Conference for Minority Students, November 09-12, 2022. Anaheim, CA

Lemas, D.J.; Lewis, B.; Frank, S.; Wright, L. Magalhães, M.; Xu, K.; Du, X.; Parker, L.; Elensi, J.; Thompson, L.; Hogan, W.R.; Modave, F. Machine learning and natural language processing for classifying infant feeding status from clinical notes. Annual Symposium of American Medical Information Association, Nov. 05-09, 2022. Washington D.C.

Du, X.; Cardel, M.I.; Millar, D.R.; Aristizabal-Henao, J.J.; Bowden, J.A.; Lemas, D.J. Untargeted urinary metabolomics analysis for an acceptance-based therapy intervention for diverse adolescent girls with overweight/obesity. ObesityWeek, Nov.1-5, 2021. Online

Du, X.; Luran, M.; Xu, K.; Kirpich, A.; Hogan, W.R.; Garrett T.J.; Lemas, D.J. A reproducible pipeline for scalable untargeted metabolomics data analysis. Annual Meeting of Metabolomics Association of North America, Sep. 14-16, 2020. Online

Lemas, D.J.; **Du, X**.; Dado-Senn, B.; Magalhães, M.; Iapicca, L.C.; Kirpich, A.; Francois, M.; Cacho, N.T.; Thompson, L.A.; Parker, L.A.; Neu, J.; Laporta, J.; Garrett, T.J. Untargeted metabolomic analysis of gestationally matched human and bovine milk sample at 2-weeks postnatal. Annual Meeting of American Society of Nutrition. May 30 – Jun 02, 2020. Online

Shah, C.; **Du, X.**; Bishnoi, R.; Bian, J. Risk of mortality in adult cancer febrile neutropenia patients with a machine learning approach. Annual Meeting of American Society of Clinical Oncology. Jun 01 – Jun 05, 2018. Chicago, Illinois

SOFTWARE AND TOOLS

RUMP. A reproducible pipeline for scalable untargeted metabolomics data analysis.

- Link: https://github.com/lemaslab/RUMP
- Role: Major contributor.

Keras. One of the most popular deep learning application programming interfaces (API) with over 375,000 users all over the world.

- Link: https://keras.io/
- Role: Contributor. My contribution enables Keras users to use Scikit-Learn to do cross-validation for deep learning models developed with Keras functional API.

PRESENTATIONS

Invited Talk. Computational reproducibility in liquid chromatography-mass spectrometry-based clinical metabolomics data processing. *University of Florida Informatics Institute*. Apr. 04, 2022. Online. [video: https://bit.ly/3O8f6Rb]

Poster Presentation. Critical review of reproducibility of LC-MS metabolomics data processing tools. *University of Florida Graduate Student Research Day*, Apr. 05, 2022. Online.

Poster Presentation. A reproducible pipeline for scalable untargeted metabolomics data analysis. *Annual Meeting of Metabolomics Association of North America*, Sep. 14-16, 2020. Online.

Conference Paper Oral Presentation. An operational deep learning pipeline for classifying life events from individual tweets. *5th International Conference on Information Management and Big Data*, Sep. 03, 2018. Lima, Peru.

TEACHING EXPERIENCE

Guest Lecturer. GMS 6804: Translational Biomedical Informatics (Dr. Dominick Lemas), University of Florida College of Medicine. Feb.-Mar. 2022. Gainesville, FL.

- Led students to visit University of Florida Research Computing Center.
- Led students to visit biomedical informatics data acquisition facilities in the University of Florida Interdisciplinary Center for Biotechnology Research.
- Introduced and discussed about research reproducibility in biomedical field, and led a discussion related to translational bioinformatics scientific papers.

Guest Lecturer. GMS 6804: Translational Bioinformatics (Dr. Dominick Lemas), University of Florida, College of Medicine. Apr. 07, 2020. Gainesville, FL.

- Introduced and discussed about a reproducible computational pipeline I developed for metabolomics data processing.

Teaching Assistant. Biomedical Informatics Summer School: Machine Learning Basics, University of Florida, College of Medicine. Jul. 23- Aug. 10, 2018. Gainesville, FL

- Taught students basic machine learning knowledges, led students to complete assignments and course projects created by myself: https://github.com/XinsongDu/Basic_ML_Practices

MENTORING EXPERIENCE

- Amanda Dobrowolski, University of Florida, 2023
  Project: Using Nextflow-based containerized workflow to process metabolomics data.

- Braeden Lewis, University of Florida, 2022
  Project: Predicting breastfeeding outcomes using machine learning approach and clinical text data.

- Ismael Wane, University of Florida, 2022
  Project: Organizing and cleaning identified metabolites using Human Metabolome Database.

- Emmanuel Elias, University of Florida, 2022
  Project: Developing a tool for automatic extraction of Human Metabolome Database ID and taxonomy information.

PROFESSIONAL TRAINING

Metabolomics Winter School
Southeast Center for Integrated Metabolomics, University of Florida, Gainesville, FL.
Jan. 27 – Jan. 29, 2020.
Description: This workshop covered cutting-edge technologies about metabolomics including sample handling, sample processing, instruments, pick picking, data analysis, etc.

Bits & Bites: Short Course Series 2021 (Online)
West Coast Metabolomics Center, UC Davis, Davis, CA.

Feb. 04 – Dec.02, 2021.
Description: This series of courses covered latest computational techniques related to metabolomics research including signal processing, metabolite annotation, statistical analysis, and data interpretation.

Focus on Mentoring Series (Online),
Office of Graduate Professional Development, University of Florida, Gainesville, FL.
Feb. 2022-Mar.2022
Description: This series of courses covered the detailed explanation of student mentoring, issues in mentoring, and ways to solve those issues. It also covered topics related to research integraty.

Coursera Certifications

- AI for Medical Diagnoses. Jun 2020
- Introduction to HTML 5. Apr 2020
- Practical Reinforcement Learning. May 2019
- Natural Language Processing. Jul 2018
- Mathematics for Machine Learning: Linear Algebra. Jun 2018
- Mathematics for Machine Learning: Multivariate Calculus. Jun 2018

PROFESSIONAL SERVICE

Reviewer for Journals:

- Journal of Big Data
- Journal of Translational Medicine
- Journal of Biomedical Informatics
- Scientific Reports
- BMC Medical Informatics and Decision Making
- BMC Pregnancy and Childbirth
- Journal of Biomolecular Techniques
- Trends in Computer Science and Information Technology

Reviewer for Conferences:

- American Medical Informatics Association (AMIA) Annual Symposium
- AMIA Summit
- AMIA Clinical Informatics Conference

Conference Program Committee:

- Service Computation 2022, Barcelona, Spain. Apr. 24 – Apr. 28, 2022 (https://www.iaria.org/conferences2022/ComSERVICECOMPUTATION22.html)

Participant of:

- The survey for producing *Times Higher Education World University Rankings*. 2021, 2022.
- The research study *Exercise as Medicine: Evaluation of a College Multidisciplinary Fitness Intervention Strategy on Perceived Wellness, Adherence, Resting Heart Rate & Blood Pressure for Sedentary Individuals*. 2022

LANGUAGES

English: Full professional proficiency

Chinese: Native proficiency

TECHNICAL SKILLS

Programming: Python, Shell, R, Groovy, Nextflow

Applications: Machine Learning, Research Software Development, High-Performance Computing, Software Containerization

Platforms: Amazon Web Services, Amazon Mechanical Turk, HiPerGator, Jupyter Notebook, Nextflow, GitHub, Docker/Singularity.

REFERENCES

Dr. Dominick J. Lemas, Ph.D., University of Florida College of Medicine, Department of Health Outcomes and Biomedical Informatics. 2004 Mowry Road-Clinical and Translational Research Building, Gainesville, FL 32610. Email: djlemas@ufl.edu. Phone: (352) 294-5971

Dr. Mathias Brochhausen, Ph.D., University of Arkansas for Medical Sciences, Department of Biomedical Informatics. 4301 West Markham Street, Little Rock, AR 72205. Email: mbrochhausen@uams.edu. Phone: (501) 686-7000

Dr. Timothy Garrett, Ph.D., University of Florida, Department of Pathology, 1395 Center Dr, Room M641c, Gainesville, FL 32610. Email: tgarrett@ufl.edu. Phone: (352) 273-5050

Dr. William R. Hogan, M.D., University of Florida College of Medicine, Department of Health Outcomes and Biomedical Informatics. 2004 Mowry Road-Clinical and Translational Research Building, Gainesville, FL 32610. Email: hoganwr@ufl.edu. Phone: (352) 294-4197