

Yu Ding
Systems Science and Industrial Engineering
Watson College of Engineering and Applied Science
Binghamton University, State University of New York
Binghamton, NY 13902
Phone: 313-329-1790
Email: yding31@binghamton.edu

EDUCATION

- **Ph.D.** Industrial and Systems Engineering Binghamton University Aug 2024(Expected)
Dissertation: Statistical Machine Learning and Data Fusion Methodologies: Applications in Healthcare (Advisor: Dr. Bing Si)
- **M.S.** Industrial Engineering Wayne State University May 2019
- **B.S.** Geophysics University of Science and Technology of China Aug 2017

HONORS AND AWARDS

- Best Paper Award
 - QCRE Best Student Paper Competition Finalist, IISE Montreal, 2024
- Grants
 - National Science Foundation (NSF) Travel Awards, 2024
 - Watson Professional Development Fund, Binghamton University, 2024
 - Conference Travel Grant 2019, University of Kansas, 2019
- Scholarships
 - University Graduate Fellowship, University of Kansas, 2019
 - Summer Research Fellowship, University of Kansas, 2019
 - Graduate Fellowship, Wayne State University, 2018
 - National Encouragement Scholarship, University of Science and Technology of China, 2016
- Kansas Half Marathon Finisher Price, Kansas, 2019

PUBLICATIONS (Published/Accepted)

1. **Ding, Y.** (2024). Statistical Machine Learning and Data Fusion Methodologies: Applications in Healthcare. *Binghamton University*. PhD Thesis
2. Sutherland, M. A., Hutchinson, M. K., Si, B., **Ding, Y.**, Liebermann, E., Connolly, S. L., ... Mueller, S. D. (2024). [Health screenings in college health centers: Variations in practice](https://doi.org/10.1080/07448481.2024.2361307). *Journal of American College Health*, 1–8. <https://doi.org/10.1080/07448481.2024.2361307>
3. Mueller, S. D. *, Sutherland, M. A., Hutchinson, M. K., Si, B., **Ding, Y.**, & Connolly, S. L. (2024). [Student Health Services at Historically Black Colleges and Universities and Predominantly Black Institutions in the United States](#). *Health Equity*, 8(1), 226-234.
4. Alramadeen, W., **Ding, Y.**, Costa, C., & Si, B. * (2023). [A novel sparse linear mixed model for multi-source mixed-frequency data fusion in telemedicine](#). *IISE Transactions on Healthcare Systems Engineering*, 13(3), 215-225. (*Selected as a Feature Article by the ISE Magazine*)

5. Jiang, L., **Ding, Y.**, Sutherland, M. A., Hutchinson, M. K., Zhang, C., & Si, B. * (2022). [A novel sparse model-based algorithm to cluster categorical data for improved health screening and public health promotion](#). *IISE Transactions on Healthcare Systems Engineering*, 12(2), 137-149.
6. **Ding, Y.**, Yang, Q.*, King, C. B., & Hong, Y. (2019). [A general accelerated destructive degradation testing model for reliability analysis](#). *IEEE Transactions on Reliability*, 68(4), 1272-1282.

PUBLICATIONS (Submitted)

7. **Ding, Y.**, Somers, V., Si, B.*. A novel sparse generalized structural equation modeling with structured sparsity for subgroup discovery from multi-modal mixed-type data. *IISE Transactions*. Under 2nd Round Review.
8. **Ding, Y.**, Costa, C., Si, B.*. Federated function-on-function regression with an efficient gradient boosting algorithm for privacy-preserving telemedicine. *IISE Transactions*. Submitted. (*Selected as QCRE Best Student Paper Competition Finalist, IISE Montreal, 2024*) [[Video](#)]
9. Mueller, S.*, **Ding, Y.**, Si, B., Sutherland, M., Hutchinson, K. Access to Campus Health Services at MSI and Non-MSI Colleges and Universities in the U.S. *Health Equity*.
10. **Ding, Y.**, Costa, C., Si, B.*. Vertical Federated Learning of Gradient Boosting for Functional Regression with Differential Privacy. *IEEE Transactions on Automation Science and Engineering*.

PUBLICATIONS (In Preparation)

11. **Ding, Y.**, Si, B.*. Functional structural equation modeling with structured sparsity for subgroup discovery. *IISE Transactions*. To be submitted.

CONFERENCE PRESENTATION

Institute of Industrial and Systems Engineers Annual Conference & Expo

- Federated function-on-function regression with an efficient gradient boosting algorithm for privacy-preserving telemedicine Montreal, May 2024

RESEARCH EXPERIENCE

My research focuses on Statistical Machine Learning and Data Fusion with Applications to Healthcare and Manufacturing Systems.

Topic 1 (Precision Medicine and Privacy-Preserving): Federated Function-on-Function Regression with an Efficient Gradient Boosting Algorithm for Privacy-Preserving Telemedicine: Precision Medicine has become a reality thanks to technological advancements in diagnostic imaging, smart sensing, and health information systems, resulting in a data-rich environment in healthcare. However, the size and complexity of data in Precision Medicine often overwhelm the modeling capabilities of existing statistical machine-learning methods. The objective of my research is to develop data fusion and system informatics approaches to enhance the quality and performance of healthcare systems, spanning screening, diagnosis, prognosis, monitoring, care, and system-level decision-making. My research proposes a pioneering federated Gradient Boosting algorithm with Least Squares Approximation (fed-GB-LSA) for efficient, privacy-preserving federated learning in the context of Function-on-Function Regression (F2R). The proposed fed-GB-LSA is

applied to health telemonitoring for Obstructive Sleep Apnea (OSA), resulting in a significant improvement in the precise diagnosis of OSA. This research aims to bridge the gap between data-rich healthcare environments and advanced machine learning, ultimately advancing the field of Precision Medicine.

Topic 2 (Diagnosis and subtyping): Multi-modal mixed-type structural equation modeling with structured sparsity for subgroup discovery from heterogeneous health data: The increasing availability of health data from large biobanks, electronic healthcare records, medical tests, wearable sensors, and more has paved the way for the development of novel machine learning (ML) models designed for multi-modal mixed-type data. These models aim to capture the complexity of human health and disease. However, clustering high-dimensional multi-modal mixed-type data poses a formidable challenge for existing ML and statistical models. To address this challenge, we propose the Multi-modal Mixed-type Structural Equation Model (M2-SEM) with structured sparsity. M2-SEM excels at uncovering precise subgroups within this intricate health data landscape, accommodating both continuous and categorical data types. We introduce a novel Gauss-Hermite-enabled Expectation-Majorization-Minimization (GH-EMM) algorithm that integrates GH quadrature and the MM algorithm within the Expectation Maximization (EM) framework for efficient model estimation. This methodology is applied to the task of identifying subgroups of individuals at low and high risk of developing adverse cardiometabolic (CM) outcomes. These outcomes encompass a full spectrum of CM risk factors, including poor nutrition, mental health, physical inactivity, and sleep deprivation.

TEACHING EXPERIENCE

Graduate Teaching Lecturer at Binghamton University (*develop course materials, give twelve lectures, host review session, and email Q&A*) Teaching Evaluation Score: 4.5/5.0

- SSIE 548: Healthcare Data Sci & Analytics (graduate class)
- ISE 448: Healthcare Data Sci & Analytics (undergraduate class)

Teaching Assistant at Wayne State University (*helped develop course materials, give lectures on selected topics and reviews, prepare exam questions and manage proctoring, and hold office hours and email Q&A*)

- IE 6430 Computer Simulation Methods (graduate class)
- IE 7270 Reliability Estimation (graduate class)

GRANT WRITING EXPERIENCE

- Helped Ph.D. advisor to prepare and conduct the following grants funded by NIH and industry:
 - NIH/NHLBI, “R01HL168173: Sleep and Cardiometabolic Subgroup Discovery and Risk Prediction in United States Adolescents and Young Adults”. Amount: \$2,452,065.
 - NIH/NHLBI, “R21HL161765: Towards Precise Phenotype Discovery of Obstructive Sleep Apnea with a Data-Inclusive Multi-Study Analysis”. Amount: \$242,770.
 - SUNY-IBM AI Research Alliance, “Ray-F2R-FL: Ray-based Functional Regression with Federated Learning”. Amount: \$200,000.
- Summer Research Fellowship. Independently wrote a proposal for Student-Led Research Grant entitled: "Operation Management Problem of Capacity Management and Innovation Investment in the Semiconductor Manufacturing Industry" presents an innovative solution to a nuanced operational challenge: striking a balance between augmenting yield rate through learning-by-doing and navigating innovation investment. This work yields actionable insights across a

spectrum of scenarios, thereby furnishing invaluable guidance to businesses that are navigating the delicate equilibrium of efficiency and patience. Amount: \$5,000.

SERVICE

- Volunteer for “Seminar Series: Statistical Modeling in Healthcare Image Pattern Recognition”, Wayne State University, 2018.

PROFESSIONAL MEMBERSHIPS

- Institute of Industrial and Systems Engineers (IISE)
- Institute for Operations Research and the Management Sciences (INFORMS)
- Institute of Electrical and Electronics Engineers (IEEE)