Tong Si (she/her)

tong.si@slu.edu|tong.si@health.slu.edu| (314)-696-9062| St. Louis, MO

Homepage: https://tongsi98.github.io/

https://www.slu.edu/medicine/health-and-clinical-outcomes-research/faculty/si-tong.php

Academic Experience

Saint Louis University Saint Louis, MO

- **Assistant Professor in the Department of Health and Clinical Outcomes Research** Jul. 2025-present

Culver-Stockton College Canton , MO

- **Assistant Professor of Computer Science & Math (tenure track)** Aug. 2024 – Jun. 2025

Saint Louis University Saint Louis, MO

- **Ph.D.** in **Statistics** (GPA 3.96/4.0) *Aug. 2020- Aug. 2024*
  Advisor: Dr. Haijun Gong
  Department of Mathematics and Statistics
- **M.S. Candidate** in **Artificial Intelligence** (GPA 3.96/4.0) *Aug. 2022-Present*
  Department of Computer Science
- **M. A**. in **Mathematics** *Aug. 2020 - May 2022*
  Department of Mathematics and Statistics

Jilin University Changchun, China

- **B. S.** in Mathematics and Applied Mathematics. *Sept. 2016 - Jun. 2020*
- **BEc in** Actuarial Science(minor) *Sept. 2017 - Jun. 2020*

Research Publications

Peer-Reviewed Papers (published)

1. Wang, Yunge, Lingling Zhang, **Tong Si**, Sarah Roberts, Yuqi Wang, and Haijun Gong. "Reconstructing Dynamic Gene Regulatory Networks Using f-Divergence from Time-Series scRNA-Seq Data." Current Issues in Molecular Biology 47, no. 6 (2025): 408.
2. Zhongyuan Zhao, Peng Zou, Yuan Fang, **Tong Si**, Bofang Yi, Tao Zhang. "Machine Learning Approaches for Assessing Medication Transfer to Human Breast Milk", *Journal of Pharmacokinetics and Pharmaco-dynamics (2025): 52:25*
3. Wen-Shan Liu, **Tong Si**, Aldas Kriauciunas, Marcus Snell, Haijun Gong, "Bidirectional f-Divergence-Based Deep Generative Method for Imputing Missing Values in Time Series Data", *Stats* 8(1), 7 (2025)
4. Yunge Wang, Lingling Zhang, **Tong Si**, Graham Bishop, Haijun Gong, "Anomaly Detection in High-Dimensional Time Series Data With a Scaled Bregman Divergence", *Algorithms* 18(2), 62 (2025)
5. **Tong Si**, Yunge Wang, Lingling Zhang, Evan Richmond, Tae-Hyuk Ahn, and Haijun Gong. "Multivariate Time Series Change-Point Detection with a Novel Pearson-like Scaled Bregman Divergence." *Stats* 7, no. 2 (2024): 462-480.
6. **Tong Si**, Zackary Hopkins, John Yanev, Jie Hou, and Haijun Gong. "A novel f-divergence based generative adversarial imputation method for scRNA-seq data analysis." *PLoS ONE* 18, no. 11 (2023): e0292792.
7. Helen Richards, Yunge Wang, **Tong Si**, Hao Zhang, and Haijun Gong. "Intelligent Learning and Verification of Biological Networks." *Advances in Artificial Intelligence, Computation, and Data Science: For Medicine and Life Science* (2021): 3-28.

Peer-Reviewed Abstracts

8. **Tong Si**, Zackary Hopkins, John Yanev, Jie Hou, and Haijun Gong. " sc-fGAIN: An f-divergence-based Generative Adversarial Imputation Method for scRNA-sq Data Analysis.", *F1000Research*, 22nd *International Conference on Bioinformatics* (2023)
9. **Tong Si,** Yunge Wang, Lingling Zhang, Kate Cannell, Haijun Gong. "Change-point detection using scaled Bregman Divergence. *F1000Research*, 22nd *International Conference on Bioinformatics* (2023)
10. **Tong Si**, Zackary Hopkins, John Yanev, Jie Hou, Haijun Gong, "f-divergence based generative adversarial imputation method for enhanced single-cell RNAseq data analysis", *International Conference on Intelligent Biology and Medicine* (2023)

Ph.D. Dissertation

11. **Tong Si**. "Missing Value Imputation and Change-Point Detection in High-Dimensional Data." PhD dissertation, Saint Louis University, (2024).

Paper Under Review

12. Noor Al Hammadi, Andrew J. Aschenbrenner, David C. Brown, Yiqi Zhu, Matthew Blake, Semere Bekena, Ramkrishna K. Singh, Johathan P. Williams, Chen Chen, **Tong Si**, Jean-Francois Trani, David B. Carr, Ramon Cassanova, Jason J. Hassenstab, Ganesh M. Babulal. "Driving Toward Early Detection of Cognitive Decline: A Novel Convolutional Neural Networks—Long Short-Term Memory Model Integrating Naturalistic Driving Data and High-Frequency Cognitive Assessments", *Submitted to NEJM AI* (2025)
13. Noor Al Hammadi, Andrew J. Aschenbrenner, David C. Brown, Yiqi Zhu, Matthew Blake, Semere Bekena, Ramkrishna K. Singh, Johathan P. Williams, Chen Chen, **Tong Si**, Jean-Francois Trani, David B. Carr, Ramon Cassanova, Jason J. Hassenstab, Ganesh M. Babulal. "Precision Detection of Cognitive Fluctuations Through Driving-Derived Digital Markers: A Deep Learning Approach" , *Submitted to The Lancet (2025).*

Research Grant

- The PhRMA Foundation Faculty Starter Grant
  Project: Generative AI for Clustering Alzheimer's and Caregiver Driving Behavior
  Role: PI; Amount: $99,828
  Starting/Ending Date: 03/01/2025 – 02/28/2026 (pending)

- The Spinal Cord Injury & Disease Research (SCIDRP) Grant (pending)
  Project: Reconstruct Time-Varying Microglial Regulatory Networks After Spinal Cord Injury Using Systems Biology Approaches
  Role: co-PI; Amount: $99,913
  Starting/Ending Date: 03/01/2025 – 02/28/2026 (pending)

- The Educational Credit Management Corporation (ECMC) Foundation
  Project: Using AI approaches and Data Analytics to Address Rural College Completion Barriers in Missouri, Illinois, and Iowa
  Role: co-PI; Amount: $228,276
  Starting/Ending Date: 2025 –2027 (pending)

Research Projects

GANs and Transformer for Separating Patient Data from Caregiver Driving Data *Sep. 2024-present*

Collaboration with Dr. Noor Al-Hammadi , Department of Health & Clinical Outcomes Research, SLU

- Developed a GAN-based clustering approach specifically tailored for large Clinical Outcomes data.
- Applied the model to extract Alzheimer's patient data from a dataset of 956,377 mixed patient and caregiver records; Expanded the approach with Transformer models to enhance data separation.
- Evaluated model performance and optimized extraction accuracy through iterative refinements.
- Preparing a manuscript for submission to a peer-reviewed journal.

Predicting PK Parameters by Antibody Glycosylation with GANs *Sep. 2024-present*

Collaboration with Dr. Tao Zhang, School of Pharmacy and Pharmaceutical Sciences, SUNY Binghamton

- Segment amino acid sequences to separate Fab and Fc regions, preparing data for analysis.
- Apply Generative Adversarial Networks (GANs) to uncover hidden relationships between antibody characteristics and pharmacokinetic (PK) parameters, enabling prediction of PK parameters.
- Reproduce and benchmark the model against existing techniques for comparison.
- Prepare a manuscript for submission to a peer-reviewed journal.

Estimating Drug Transfer into Breast Milk using Machine Learning Approach *Oct. 2023-Sep. 2024*

Collaboration with Dr. Tao Zhang's lab, School of Pharmacy and Pharmaceutical Sciences, SUNY Binghamton

- Categorize Milk/Plasma (M/P) drug ratios into distinct intervals based on a comprehensive literature review.
- Implement various machine learning models, including K-Nearest Neighbors (KNN), Random Forest, Support Vector Machine (SVM), and Deep Neural Networks in Python.
- Compare model performance, providing novel insights into the prediction of M/P ratios.
- Submitted a manuscript [7] to a peer-reviewed journal, currently under review.

Imputation of Time Series Data via Generative Models and GRU *Oct. 2023-Present*

Team leader, Dr. Gong's group, SLU

- Conduct a thorough literature survey on time series data imputation to identify prevalent limitations and gaps in current methodologies.
- Develop a GRU-based time-series generative adversarial imputation network algorithm and investigate the mathematical theory underlying the algorithm; Implement the time series imputation algorithm based on different divergence functions using Python.
- **A paper [2] was published in Stats in 2025**, I am a **corresponding author**.

Change-Point Detection for Time Series Data Using Scaled Bregman Divergence *June 2023 - Dec 2024*

Team leader, Dr. Gong's group, SLU

- Developed a Pearson-like Scaled Bregman Divergence Method [1,3] for Change-point Detection [3] and Anomaly detection [1] in multivariate time series data; investigated the mathematical foundation of the algorithm and reinforce the algorithm's generality and reliability across a broader range of applications.
- Reproduced comparative methods in R and Python to benchmark our model against existing techniques. Compare the accuracy in identifying change-points, and performance across diverse datasets and conditions.
- **Two papers [1,3] were published in *Stats* in 2024 and *Algorithms* in 2025**. I am a **corresponding author** in [3].

Imputation of sc-RNA Sequencing Data via Generative Adversarial Networks *Oct. 2022 -May. 2023*

Team leader, Dr. Gong's group, SLU

- Led the team to develop a novel single cell f-divergence based generative adversarial imputation network (sc-fGAIN) algorithm to impute the missing values in the single cell RNA sequencing data.
- Implemented the sc-fGAIN algorithm using Python and provide mathematical proofs to confirm its effectiveness and general applicability in imputation tasks.
- Managed a massive dataset with dimensions 10,164 by 3,918, ensuring efficient data preprocessing and algorithm application.
- Implemented and compared different state-of-the-art imputation methods as benchmarks using R, Python, and MATLAB to validate the superiority of our approach.
- **A paper [4] was published in PLOS ONE in 2023**. I received a **Best Oral Presentation Award** [8] at 2023 **International Conference on Bioinformatics**, held in Brisbane, Australia.

Innovative Web-Based Library Management System *Sept. 2023 - Dec. 2023*

Team leader of Course Project, SLU

- Utilized SQL for robust database design and management, ensuring efficient data storage, retrieval, and manipulation; Implement the user interface using HTML, creating an intuitive and responsive web application.
- Built the core functionality of the system using Python, ensuring seamless integration with the database and frontend components;
- Apply GitHub for source code management and team collaboration, maintaining an organized and efficient development workflow.
- Used CircleCI for continuous integration, automating code testing and deployment processes, to enhance code quality and deployment efficiency
- Employed Docker Hub for containerizing the application, ensuring consistent deployment across different environments.

Analytical Text Processing Using Machine Learning *Sept. 2022 - Nov. 2022*

Course Project, SLU

- Applied Python libraries Pandas for data manipulation and Scikit-Learn for machine learning model implementation, including using feature sklearn.feature_extraction.text.CountVectorizer for text preprocessing and feature extraction
- Processed raw text data using tokenization and lemmatization techniques.
- Implemented a variety of classification algorithms, including Naive Bayes, SVM, and Random Forest, to compare performance. Optimize models using cross-validation and grid search techniques.

Statistical Inference and verification of Regulatory Networks *Sept. 2020 - May 2021*

Collaborative Research Project, Dr. Gong's group, SLU

- Applied a weighted dynamic Bayesian network method to reconstruct gene regulatory network from time series microarray data with other team members.
- Implemented different model checking technique, including SMV and PRISM for the network verification.
- **A paper [5] was published in 2021.**

Teaching at Culver-Stockton College

Computer Programming (Python) *Fall 2024*
Elementary Statistics *Fall 2024*
Beginning Math *Fall 2024*
College Algebra *Fall 2024*
Calculus II Spring 2025
Applications of Python Programming Spring 2025

Teaching at Saint Louis University

Instructor of College Algebra *Jan. 2022-Dec. 2022*
Teaching Assistant: Regression Analysis; Bayesian Statistics & Statistical Computing *2023-2024*
Teaching Assistant: Calculus I *Aug. 2021 – Dec. 2021*

Professional Service

NSF Proposal Panel: ad hoc Reviewer *Nov. 2024-present*
Topical Advisory Panel (TAP) for *Current Issues in Molecular Biology*, MDPI *Sep.2024-present*
Member of International Society for Computational Biology *Mar. 2024-present*
Research Assistant, Dr. Gong's Group, Saint Louis University *Jan. 2023-Present*
Reviewer of the following Journals: BMC Bioinformatics; Heliyon; PLOS ONE ; Journal of Bioinformatics & Computational Biology; Journal of Theoretical Biology; Genomics …
Treasurer of Association for Women in Mathematics (AWM), SLU *Aug. 2022- Jan. 2023*

Conference Presentation

Oral Presentation at **International Symposium on Bioinformatics Research and Applications** Jul. 2024
Poster Presentation **at 16th Great Lakes Bioinformatics (GLBIO) conference, Pittsburgh, PA** *May.2024*
Oral Presentation **at Annual Graduate Research Symposium, Saint Louis University** *Apr. 2024*
Oral Presentation **at the Mathematical Association of America Missouri Section, Liberty, MO** Apr. 2024
Oral Presentation **at the Danforth Plant Sciences Center, St. Louis, MO** Jan. 2024
Oral Presentation, **22nd International Conference on Bioinformatics, Brisbane, Australia** *Nov. 2023*
Poster Presentation, **International Conference on Intelligent Biology & Medicine, Tampa, FL** *Jul. 2023*

Awards and Certificate

Best Oral Presentation Award, 22nd International Conference on Bioinformatics, Australia *Nov. 2023*
Full financial support for Mathematical Problems in Industry (MPI) Workshop *Jun. 2024*
Full financial support for Graduate Student Mathematical Modeling Camp (GSMMC) *Jun. 2024*
2nd Place for Oral Presentation Award, Annual Graduate Research Symposium, SLU *Apr. 2024*
GLBIO 2024 travel fellowship *Mar.2024*
Dean's Travel Award, Saint Louis University *2023-2024*

Travel Award, Forty Third Midwest Probability Colloquium *Oct. 2022*

Teaching Certificate, Saint Louis University *Sept. 2025*

Certificate of Data Science with SQL and Tableau from Cornell University *Feb. 2025*

SKILLS AND CERTIFICATIONS

- Computer Skills: Python, R, MATLAB, SQL, HTML, Tensorflow, Pytorch, HPC
- Skills: Data analysis for big data, Software development, Database skills, Website building skills