

# Shengpu Tang

## Curriculum Vitae

2260 Hayward St, Beyster 3861  
Ann Arbor, MI 48109, USA  
+1 (734) 834 7835  
tangsp@umich.edu  
shengpu-tang.me  
shengpu1126

### Education

#### Ph.D. in Computer Science & Engineering.

University of Michigan, Ann Arbor

Advisor: Jenna Wiens.

Present

Expected graduation date: May 2023

#### M.S.E. in Computer Science & Engineering

University of Michigan, Ann Arbor

April 2020

GPA: 3.938/4

#### B.S.E. in Computer Science

University of Michigan, Ann Arbor

Summa Cum Laude. Minor in Mathematics.

April 2018

GPA: 3.972/4

### Research Interests

My research focuses on developing and applying machine learning methods to help solve important problems related to healthcare. While motivated by specific use cases (such as risk stratification and dynamic treatment recommendation for various diseases and medical conditions), my work aims to identify interesting technical challenges and propose novel solutions that are broadly relevant to fundamental AI/ML research. My past contributions span the technical areas of *reinforcement learning*, *deep learning*, *time-series/sequential data modeling*, *transfer learning*, *causality*, *algorithmic fairness*, etc., and I hope to apply my knowledge and expertise to interdisciplinary problems in healthcare and beyond.

### Skills

Computing Python, C, C++, C#, MATLAB, Swift, HTML5, JavaScript, React, numpy, pandas, sklearn, matplotlib, seaborn, joblib, PyTorch, tensorflow, Keras, jupyter, Visual Studio Code, Unity3D.

Language English (professional), Chinese Mandarin (native)

### Publications

Note: \* and † denote equal contributions.

#### Conferences

##### C3. Model Selection for Offline Reinforcement Learning: Practical Considerations for Healthcare Settings

Shengpu Tang & Jenna Wiens

*Machine Learning for Healthcare Conference (MLHC 2021)*

Online. Aug 2021. Acceptance rate: 37.3% (31/83).

[code] [paper]

► Also at *RL4RealLife Workshop @ ICML 2021*

C2. **Clinician-in-the-loop Decision Making: Reinforcement Learning with Near-Optimal Set-Valued Policies**

Shengpu Tang, Aditya Modi, Michael W. Sjoding, & Jenna Wiens  
*International Conference on Machine Learning (ICML 2020)*

Online. July 2020. Acceptance rate: 21.8% (1,088/4,990).

[code] [paper]

C1. **Relaxed Parameter Sharing: Effectively Modeling Time-Varying Relationships in Clinical Time-Series**

Jeeheh Oh\*, Jiaxuan Wang\*, Shengpu Tang, Michael W. Sjoding, & Jenna Wiens  
*Machine Learning for Healthcare Conference (MLHC 2019)*

Ann Arbor, MI. Aug 2019. Acceptance rate: 30.9% (38/123).

[code] [paper]

---

**Journal Articles**

J4. **Predicting Postoperative Opioid Use with Machine Learning and Insurance Claims in Opioid-Naïve Patients**

Jaewon Hur\*, Shengpu Tang\*, Vidhya Gunaseelan, Joceline Vu, Chad M. Brummet, Michael Englesbe, Jennifer Waljee†, Jenna Wiens†  
*The American Journal of Surgery.*

Published online Mar 26, 2021.

[code] [paper]

J3. **Evaluating a Widely Implemented Proprietary Deterioration Index Model Among Hospitalized COVID-19 Patients**

Karandeep Singh, Thomas S. Valley, Shengpu Tang, Benjamin Y. Li, Fahad Kamran, Michael W. Sjoding, Jenna Wiens, Erkin Otles, John P. Donnelly, Melissa Y. Wei, Jonathon P. McBride, Jie Cao, Carleen Penozza, John Z. Ayanian, Brahmajee K. Nallamothu  
*Annals of the American Thoracic Society*, 2021 Jul;18(7):1129-1137.

Published online Dec 23, 2020.

[code] [paper]

J2. **Democratizing EHR Analyses with FIDDLE – A Flexible Data-Driven Preprocessing Pipeline for Structured Clinical Data**

Shengpu Tang, Parmida Davarmanesh, Yanmeng Song, Danai Koutra, Michael W. Sjoding, Jenna Wiens

*Journal of the American Medical Informatics Association*, 2020;27(12):1921-1934.

Published online Oct 11, 2020.

[code] [paper]

► Accompanying Dataset: **MIMIC-III and eICU-CRD: Feature Representation by FIDDLE Preprocessing.**  
*PhysioNet*. Available online, April 28, 2021. [link]

► Previously presented at MLHC 2019 as a 1-page clinical abstract. [link]

J1. **Predicting Acute Graft-vs-Host Disease using Machine Learning and Longitudinal Vital Sign Data in Electronic Health Records**

Shengpu Tang, Grant T. Chappell, Amanda Mazzoli, Muneesh Tewari, Sung Won Choi\*, and Jenna Wiens\*

*JCO Clinical Cancer Informatics*, 4(2020):128-135.

Published online Feb 21, 2020.

[code] [paper]

## Preprints/Work-in-Progress

- **Development and Multi-Site External Validation of an Open Source Real-Time Deterioration Index**  
Fahad Kamran\*, Shengpu Tang\*, Erkin Otles, Dustin McEvoy, Sameh Saleh, Jen Gong, Benjamin Y. Li, Sayon Dutta, Xinran Liu, Richard Medford, Thomas S. Valley, Lauren R. West, Karandeep Singh, Seth Blumberg, John Donnelly, Erica Shenoy, John Ayanian, Brahmajee K. Nallamothu, Michael W. Sjoding†, Jenna Wiens†  
In preparation, 2021.
- **Respecting Autonomy and Enabling Diversity: the effects of “neutral” recruitment and consent methods on the diversity of research databanks**  
Kayte Spector-Bagdady\*, Shengpu Tang\*, Sarah Jabbour\*, W. Nicholson Price II, Ana Bracic, Sachin Kheterpal, Chad M. Brummett, Jenna Wiens  
Under review, 2021.
- **A BioHackathon on the challenges of aging: Rapid need-finding and solution modeling in a student-led event**  
James Ashton-Miller, Loubna Baroudi, Barry Belmont, Natacha Comandante-Lou, Jesus A. Castor-Macias, Kathleen E. Finn, Dorsa Haji Ghaffari, Robert Graham, Jacqueline Larouche, Benjamin Y. Li, Ying Liu, Anjali Mittal, Zachariah Sperry, Shengpu Tang, Nathaly Villacis, Megan Weivoda, Raymond Yung, Hans J. Zander, Yingying Zeng, Muru Zhou  
engrXiv, May 30, 2019 [paper]

## Talks & Presentations

5. *Reinforcement Learning with Set-Valued Policies: A Framework for Clinician-in-the-Loop Decision Making*. PathCheck Global Health Innovators Seminar, Aug 12, 2021. 15-min [Video] (+Q&A)
4. *Model Selection for Offline Reinforcement Learning: Practical Considerations for Healthcare Settings*. MLHC 2021. 3-min [Video]
3. *Clinician-in-the-loop Decision Making: Reinforcement Learning with Near-Optimal Set-Valued Policies*. ICML 2020. 15-min [Video]
2. *FIDDLE: A Comprehensive, Generalizable Pipeline for Learning from Clinical Data*. MiCHAMP Seminar Series, Jan 10, 2020. (with Parmida Davarmanesh) 45-min [Slides]
1. *Relaxed Parameter Sharing: Effectively Modeling Time-Varying Relationships in Clinical Time-Series*. MLHC 2019. 3-min [Slides]

## Posters

7. *Model Selection for Offline Reinforcement Learning: Practical Considerations for Healthcare Settings*. MLHC 2021. [Poster]
6. *Model Selection for Offline Reinforcement Learning: Practical Considerations for Healthcare Settings*. RL4RealLife Workshop @ ICML 2021.
5. *Clinician-in-the-loop Decision Making: Reinforcement Learning with Near-Optimal Set-Valued Policies*. Michigan AI Symposium 2020. [Poster]
4. *Democratizing EHR Analyses – A Comprehensive, Generalizable Pipeline for Learning from Clinical Data*. Michigan AI Symposium 2019.
3. *Democratizing EHR Analyses – A Comprehensive, Generalizable Pipeline for Learning from Clinical Data*. MLHC 2019. [Poster]
2. *Relaxed Parameter Sharing: Effectively Modeling Time-Varying Relationships in Clinical Time-Series*. MLHC 2019. [Poster]
1. *Comparing Deep and Shallow Techniques for Building Risk Stratification Models using EHR Data*. MIDAS Annual Symposium 2018. [Poster]

---

## Professional Service

---

### Reviewer

NeurIPS | 2020<sup>†</sup>, 2021.

NeurIPS Workshops | 2021.

ICML | 2021<sup>§</sup>.

ICLR | 2022.

MLHC | 2020, 2021.

RL4RealLife Workshop @ ICML | 2021.

<sup>†</sup>Top reviewer award, <sup>§</sup>Emergency/Expert reviewer.

---

### Organizer

- **Bridging the Gap: from Machine Learning Research to Clinical Practice**  
NeurIPS 2021 Workshop, Virtual Dec 2021
  - Co-organized the workshop with: Melanie Fernandez, Sonali Parbhoo, Patrick Schwab, Julia Vogt, Mario Wieser, Jiayu Yao.
  - Focused around the topic: How can we enable the use of cutting-edge machine learning research in daily clinical practice and ultimately improve patient care?
  - Managed invited speakers, workshop website, paper reviewing, spotlight talks, poster session, panel discussion.
- **MLHC 2019 | Community Data Challenge**  
University of Michigan, Ann Arbor ( $\sim$ 120 participants) Aug 2019
  - Facilitated the first-ever datathon-style community challenge for MLHC, together with: Brahmajee Nallamothu, Jenna Wiens, Erin Kaleba, Michael W. Sjoding, John Vandervest, Tom Pollard, Alistair Johnson, Omar Badawi.
  - Guided participants to work in interdisciplinary teams and formulate technically interesting and clinically relevant problems. The database and notebooks were hosted through Google Cloud Platform BigQuery and Datalab. [link]

---

### Panelist

#### University of Michigan, Division of Computer Science & Engineering (UM-CSE)

- Explore Graduate Studies Oct 2019, Oct 2020
- Orientation for Incoming CSE Master's Students Sept 2020

---

## Teaching & Mentoring

- **Summer Undergraduate Research in Engineering 2020**  
College of Engineering, University of Michigan, Ann Arbor. May 2020 - Aug 2020  
*Research Mentor* (2 students)
  - Guided undergraduate students through summer machine learning projects to explore the following research questions:
    - (i) Can we extend Sequence Transformer Networks (STN) to incorporate local and feature-specific transformations?
    - (ii) Does STN outperform other deep-learning models on the UM Acute Dyspnea dataset?
  - Helped students prepare 1-page abstracts as deliverable at the end of the summer.

- **Big Data Summer Institute 2019**  
School of Public Health, University of Michigan, Ann Arbor June 2019 - July 2019  
*Research Mentor* (9 students)
  - Guided undergraduate students to work on the task of predicting in-hospital mortality in ICU patients (The PhysioNet/CinC Challenge 2012) to explore the following research questions:
    - (i) Can we augment the standard 1D-CNN architecture by incorporating a reliability metric that accounts for irregularities and missingness in the measurement patterns?
    - (ii) How robust are CNN and LSTM against random and adversarial noise?
    - (iii) Do different models (LR, RF, CNN, LSTM) with similar discriminative performance yield similar explanations of their predictions?
  - Helped students prepare posters and presentations at a concluding research symposium.
- **EECS 445 Introduction to Machine Learning**  
College of Engineering, University of Michigan, Ann Arbor Sept 2018 - Dec 2018  
*Graduate Student Instructor* (211 students)
  - Wrote homework questions, led discussion sessions, held office hours, managed grading.
  - Significantly redesigned the course assignments, specifically:
    - (i) Introduced The PhysioNet/CinC Challenge 2012 dataset for ICU patient mortality prediction using clinical data in the first project;
    - (ii) Rewrote the entire assignment in PyTorch (originally in TensorFlow 1.x) for image classification in the second project;
    - (iii) Added a homework problem on LSTM for sequence classification.
  - Student rating: 4.8/5 (36 students in discussion section).
- **ESSI Summer Camp: Unsupervised Learning and Clustering**  
Exercise and Sports Science Initiative (ESSI), University of Michigan, Ann Arbor June 2018  
*Guest Lecturer* (~20 high school students)
  - Gave a lecture on unsupervised learning.
  - Led a hands-on tutorial on clustering of NBA players. [Exercise] [Solution]
- **Big Data Summer Institute 2018**  
School of Public Health, University of Michigan, Ann Arbor June 2018 - July 2018  
*Research Mentor* (15 students)
  - Guided students to conduct research on the topic of predicting acute respiratory failure in ICU patients (MIMIC-III) using: (i) 1D-CNN, (ii) LSTM, (iii) interpretable models (LR & RF).
  - Helped students prepare posters and presentations at a concluding research symposium.
- **EECS 445 Introduction to Machine Learning**  
College of Engineering, University of Michigan, Ann Arbor Sept 2017 - Dec 2017  
*Instructional Aide* (124 students)
  - Wrote homework questions, led discussion sessions, held office hours, managed grading.
- **Math Lab**  
Department of Mathematics, University of Michigan, Ann Arbor Sept 2016 - Apr 2017  
*Tutor*
  - Provided walk-in tutoring service to other students for various college mathematics courses, ranging up to linear algebra, advanced calculus, and differential equations.

## Honors & Awards

|   |           |
|---|-----------|
| <b>James B. Angell Scholar.</b> University of Michigan, Ann Arbor | 2017-2019 |
| <b>University Honors.</b> University of Michigan, Ann Arbor       | 2015-2018 |
| <b>Dean's List.</b> University of Michigan, Ann Arbor             | 2015-2018 |

## Mentorship

|   |      |
|---|------|
| <b>Mentor for Undergraduate Researcher</b><br>Allen Diao  | 2021 |
| <b>Summer Undergraduate Research in Engineering (SURE)</b><br>Allen Diao, Brian Lin   | 2020 |
| <b>Big Data Summer Institute (BDSI)</b><br>Yujuan Fu, Michael Wieck-Sosa, Diego Moreno, Eric Chen, Samuel Hawk, Thea Sukianto, Kyla E. Chasalow, Wen Si, Andrew Brian Zehr  | 2019 |
| <b>Big Data Summer Institute (BDSI)</b><br>Saptarshi Chakraborty, Un Young Chung, Parmida Davarmanesh, Conrad Gordon, Xin Gu, Gabriel Kralik, Sophia Luo, Zeltzyn Montes Rosales, Swaraj Nayegandhi, Zhu Shen, Yanmeng Song, Sean Tocci, Eric Yanchenko | 2018 |

## Open Source Projects

- pandas v0.24.0: Bug in `DatetimeIndex` subtraction that incorrectly failed to raise `OverflowError` (GH22492, GH22508)

## Other Industry Experience

- **LendingHome**. San Francisco, CA. May 2017 - Aug 2017  
*Software Engineering Intern*
  - Worked on website backend/frontend using Ruby on Rails and React JS.
  - Developed and incorporated conditions template into the project workflow.