Yiyang Sun

(919) 667 - 6375 | email: ys250@duke.edu | williamsyy.github.io

EDUCATION

Duke University ***Aug 2024 – Present***
PhD in Computer Science ***Supervisor: Cynthia Rudin*** ***Cum GPA: 4.0/4.0***

Duke University ***Sep 2022 – May 2024***
Master of Science: Electrical and Computer Engineering (Full Scholarship) ***Cum GPA:3.93/4.0***

Duke Kunshan University (DKU) / Duke University ***Aug 2018 – May 2022***
Bachelor of Science: Data Science ***Cum GPA:3.82/4.0***

RESEARCH INTERESTS

Interpretable Machine Learning, Interpretability and Explainability, Interpretable Representation Learning

PROJECT EXPERIENCE

Interpretable Machine Learning Lab, Duke University

Dictionary Learning for LLM Interpretability ***Nov 2024 – Present***
- Decompose the concepts for latent space representations to improve the mechanistic interpretability for LLM structures using sparse autoencoders

Dimension Reduction ***Mar 2024 – Present***
- Develop algorithms for improving the embedding space for real and stable cluster separation
- Work on the improvement of parametric-based dimension reduction techniques

Sparse Explanation Values ***Mar 2023 – Present***
- Define Sparse Explanation Value (SEV) that measures prediction explanation sparsity.
- Prove that most predictions are inherently sparse, even if models are not inherently sparse.
- Develop new regularization terms that lead to sparser explanations while maintaining model performance.

Sustainable Expedited Algae plastic Discovery (SEAD) Team, Duke University

Interpretable Feature Representation and Machine Learning Model Construction ***Jan 2024 – Present***
- Extract useful features from the high-throughput data from the material experiments
- Develop surrogate models for predicting material microstructure to properties
- Develop optimization process for estimating the best processing conditions through inverse design techniques

iMEP Research Center, DKU & University of Georgia [joint research]

Interpretable representation learning for single-cell RNA data ***Apr 2022 – May 2022***
- Set up benchmarks for various single cell embedding techniques
- Build up interpretable deep learning models based on a gene regulatory network-guided framework.

Estimating catalytic constant based on protein structure ***Mar 2022 – May 2022***
- Develop a catalytic constant estimation heterogeneous attention graphical neural network based on interactions and enzyme protein structures embedded in AlphaFold 2

DKU & Shanxi Hospital [joint research]

Diabetes Kidney Diseases Analysis ***Mar 2021 – May 2022***
- Analyzed Diabetes Kidney Diseases (DKD) complications
- Created Non-DKD diagnosis models that avoided kidney puncture operations for DKD patients
- Estimate the risk of DKD and NDKD in Type 2 Diabetes Mellitus patients based on medicine intake

Stroke Factor Analysis ***Sept 2020 – Aug 2021***
- Performed causal analysis for stroke patients with interpretable machine learning algorithms.
- Developed an interpretable machine learning model in identifying main stroke for stroke.

PUBLISHED PAPERS

[1] Y. Wang [1], **Y. Sun [1]**, H. Huang [1], C. Rudin. **Dimension Reduction with Locally Adjusted Graphs**, AAAI 2025

[2] **Y. Sun**, T. Wang, C. Rudin. **Improving Decision Sparsity**. *NeurIPS 2024*. https://doi.org/10.48550/arXiv.2410.20483

[3] **Y. Sun** [1], Z. Chen [1], V. Orlandi, T. Wang, C. Rudin. **Sparse and Faithful Explanations Without Sparse Models**. ***Data Mining Best Paper***, *INFORMS 2023*, *AISTATS* 2024. doi: https://doi.org/10.48550/arXiv.2402.09702

[4] D. Hui [1], **Y. Sun** [1], S. Xu, et. Al. **Analysis of clinical predictors of kidney diseases in type 2 diabetes patients based on machine learning**, *International urology and nephrology 55.3 (2023): 687-696.* doi: https://doi.org/10.1007/s11255-

022-03322-1(2023)

[5] J, Ma, **Y. Sun**, J. Liu, et al. **Multi-objective learning and explanation for stroke risk assessment in Shanxi province.** *Science Reports*,12,22337. doi: https://doi.org/10.1038/s41598-022-26595-z (2022)

[6] J Liu [1], **Y. Sun** [1], J Ma, et al. **Analysis of main risk factors causing stroke in Shanxi Province based on machine learning models**. *Informatics in Medicine Unlocked*,26(5):100712. https://doi.org/10.1016/j.imu.2021.100712 (2021)

[7] Q. Lu, Y. Xiong, **Y. Sun**, et al. **Comparative experimental study on the mechanical properties of red blood cells from newborn babies and elderly people**. *Journal of medical biomechanics*, 36(04):638-645 https://doi.org/10.16156/j.1004-7220.2021.04.021 (2021)

RESEARCH EXPERIENCE

Lab Member | *Duke University* — *Durham, United States*
Interpretable Machine Learning Lab | Advisor: Cynthia Rudin — **Mar 2023 – Present**

Research Assistant | *DKU* — *Kunshan, China*
Center for Mathematics and Computational Sciences | Advisor: Shixin Xu, Huaxiong Huang — **Mar 2020 – May 2022**

Research Assistant | *DKU* — *Kunshan, China*
iMEP Research Center | Advisor: Huansheng Cao, Gaoyang Li — **Mar 2021 – May 2022**

Lab Member | *University of Toronto* — *Toronto, Canada*
2020 Fields Undergraduate Summer Research Program | Advisor: Huaxiong Huang — **Jul 2020 – Aug 2020**

Research Assistant | *DKU* — *Kunshan, China*
Global Health Research Center | Advisor: Lijing Yan — **Mar 2019 – Jun 2020**

WORK EXPERIENCE

Teaching Assistant | *Duke University* — *Durham, United States*
Probailistic Machine Learning (STA561) — **Jan 2025 – Present**
Theory and Algorithm of Machine Learning (CS671) — **Aug 2023 – Dec 2023**
- Design assignment questions, hold office hours and lead discussions.

President | *DKU* — *Kunshan, China*
Mathematical Modeling Club — **Mar 2019 – May 2022**
- Set up introductory programming classes and held weekly experience-sharing sessions
- Invited professors from various fields to give lectures on the application of math modeling
- Organized school-level activities and competitions

Experiment Analyst Intern | *Merck millipore* — *Shanghai, China*
MLab — **Jul 2020 – Aug 2020**
- Conducted experiments of Chinese Hamster Ovary (CHO) cells under different growing conditions
- Performed solution turbidity analysis based on filters and industrial operation processes

Academic Service

Top reviewer in NeurIPS2024, Reviewer in ICLR2025, AISTATS2025, TMLR 2025,

AWARDS

AI+Material Fellowship — **Aug 2024**
aiM Program — *Duke University*

Outstanding Graduate Teaching Assistant Award — **May 2024**
Pratt School of Engineering — *Duke University*

Best Paper Award — **Oct 2023**
INFORMS 2023 Data Mining Competition (General Track) — *INFORMS Annual Meeting*

Honorable Mention — **Apr 2021**
2021 MCM/ICM Contest — *COMAP*

Math Innovation Award — **Mar 2021**
DKU

First Prize — **Jul 2020**
Mathorcup National Mathematical Modeling Contest — *Operations Research Society of China*