

# Huan He

☎ (+1) 860-617-6686  
✉ Huan.He@Pennmedicine.upenn.edu  
📄 hehuannb.github.io

My name is Huan He, a postdoctoral researcher at University of Pennsylvania, working with Professor Yong Chen. I received my PhD in Computer Science from Emory in 2022, where I was co-supervised by Professors Joyce Ho and Yuanzhe Xi.

A main thrust of my research focuses on develop efficient and trustworthy machine learning models for complex real-world data (e.g., tensors, images, texts, graphs, RNA-Seq), under different contexts (e.g., distribution shift, missing data), from various aspects (e.g., accuracy, efficiency, robustness, fairness). Thus far, I have focused on two complementary directions: (1) **Scientific Computing for Machine Learning** (e.g., optimization, numerical analysis, parallel computing) (2) **Machine Learning for Health** (e.g., Electronic Health Records Phenotyping, Causal Inference, Transfer Learning)

## Professional Experience

- 2023-Present **Postdoc**, *University of Pennsylvania*, Advisor: Dr. Yong Chen.
- 2022-2023 **Postdoc**, *Harvard Medical School*, Advisor: Dr. Marinka Zitnik.

## Education

- 2016-2022 **Ph.D in Computer Science**, *Emory University*, Advisor: Dr. Joyce C. Ho and Yuanzhe Xi.
- 2014-2016 **M.Sc in Financial Mathematics**, *University of Connecticut*, Advisor: Dr. James G. Bridgeman.
- 2010-2014 **B.S in Financial Engineering**, *Shanghai Finance University*.

## Publications and Preprints

- 2024 ICLR **Huan He**, William hao, Yuanzhe Xi, Yong Chen, Bradley Malin, Joyce Ho. A Flexible Generative Model for Heterogeneous Tabular EHR with Missing Modality. In: *International Conference on Learning Representations (ICLR) 2024*.
- 2023 Xiaokang Liu, **Huan He**, Naimin Jing, Jason Moore, Christopher Forrest, Yong Chen. PreTrained-Subtyping: Pre-trained knowledge-guided transfer learning for identifying clinical subphenotypes of multisystem inflammatory syndrome in children. *Under Review*
- 2023 Ru Huang, Kai Chang, **Huan He**, Ruipeng Li, Yuanzhe Xi. Reducing operator complexity in Algebraic Multigrid with Machine Learning Approaches. *Minor Revision(Revised)*, *SIAM Journal on Scientific Computing*
- 2023 **Huan He**, Shifan Zhao, Ziyuan Tang, Yousef Saad, Yuanzhe Xi. NLTGCR: A class of Nonlinear Acceleration Procedures based on Conjugate Residuals. *Accepted by SIAM Journal on Matrix Analysis and Applications (SIMAX)*.
- 2023 Owen Queen, Thomas Hartvigsen, Teddy Koker, **Huan He**, Theodoros Tsiligkaridis, Marinka Zitnik: Encoding Time-Series Explanations through Self-Supervised Model Behavior Consistency. In: *Conference on Neural Information Processing Systems (NeurIPS 2023, Spotlight)*
- 2023 **Huan He**, Owen Queen, Teddy Koker, Consuelo Cuevas, Theodoros Tsiligkaridis, Marinka Zitnik. Domain Adaptation for Time Series Under Feature and Label Shifts. In: *International Conference on Machine Learning (ICML) 2023*.
- 2023 Jiali Cheng, **Huan He**<sup>\*</sup>, George Dasoulas<sup>\*</sup>, Chirag Agarwal, Marinka Zitnik. GN-Delete: A General Unlearning Strategy for Graph Neural Networks. (<sup>\*</sup>: equal contribution). In: *International Conference on Learning Representations (ICLR) 2023*.

- 2023 **Huan He**, Joyce C Ho, Shifan Zhao, Yuanzhe Xi. MedDiff: Denoising Diffusion Probabilistic Models for Electronic Health Records generation . *In Review*.
- 2022 Yuliang Ji, Difeng Cai, **Huan He**, Yuanzhe Xi. AUTM Flow: Atomic Unrestricted Time Machine for Monotonic Normalizing Flows. In: *Conference on Uncertainty in Artificial Intelligence (UAI) 2022*.
- 2022 **Huan He**, Shifan Zhao, Yuanzhe Xi, Joyce C Ho, Yousef Saad. GDA-AM: Solve Minimax Optimizatlon by Anderson Mixing. In: *International Conference on Learning Representations (ICLR) 2022*.
- 2021 **Huan He**, Yuanzhe Xi, Joyce C Ho. AGE: Enhancing the Convergence on GANs using Alternating extra-gradient with Nonlinear Gradient Extrapolation. In: *NeurIPS Workshop on Deep Generative Models and Downstream Applications (Oral)*, 2021.
- 2020 **Huan He**, Yuanzhe Xi, Joyce C Ho. Fast and Accurate Tensor Decomposition without a High-Performance Computing Machine. In: *International Conference on BigData*. 2020.
- 2020 **Huan He**, Yuanzhe Xi, Joyce C Ho. Accelerated Stochastic Gradient Decent(SGD) for Sparse Tensor Decomposition. In: *IEEE International Conference on Data Mining Workshop*, 2020.
- 2019 **Huan He**, Jette Henderson, Joyce C Ho. SGranite: Distributed tensor decomposition for large scale health analytics. In: *Proceedings of The Web Conference*. 2019.
- 2018 Jette Henderson, **Huan He**, Bradley A Malin, Joshua C Denny, Abel N Kho, Joydeep Ghosh, and Joyce C Ho. Phenotyping through Semi-Supervised Tensor Factorization (PSST). In: *AMIA Annual Symposium*. 2018.

## High-level Research Projects

### 2018-2023 Scientific Computing for Efficient Machine Learning .

- **Nonlinear Acceleration Methods and Optimization Techniques:**
  - Focused on optimization methods for complex problems
  - Developed a nonlinear Truncated Generalized Conjugate Residual method (nlTGCR) to reduce memory usage by exploiting the symmetry of the Hessian
  - Developed the **first work** that solves the diverging or cycling behavior of gradient descent-ascent on minimax optimization problems by tapping into Anderson Acceleration
  - Proved accelerated convergence rates and demonstrated improved convergence compared to SOTA methods on classical and deep learning problems
- **Parallel Computing:**
  - Designed smart parallel computing algorithms to enhance the performance of machine learning models training
  - Achieved up to  $5\times$  speedup compared with the state-of-the-art GAN models
  - Achieved at least a  $4\times$  speed-up compared to a state-of-the-art distributed tensor factorization method
- **Transfer Learning:**
  - Focused on robust machine learning deployment for data-scarcity scenarios
  - Developed high-accuracy out-of-distribution data models using domain adaptation
  - Showed significant improvement on immunotherapy response prediction and time series classification
- **Generative Models with Enhanced Convergence**
  - Focus on training generative models with enhanced convergence
  - Developed the **first work** that solves the diverging or cycling behavior of GDA on minimax optimization problems by tapping into Anderson Acceleration
  - Developed a normalizing flow that is able to learn both latent feature and tractable marginal likelihood estimation
  - Developed diffusion models to generate high-fidelity synthetic electronic health records

2018-2023 **Healthcare.**

○ **Longitudinal Electronic Health Records Phenotyping :**

- Developed a semi-supervised tensor decomposition method that innovatively combines labeled and unlabeled data, enhancing the accuracy and relevance of patient phenotypes extracted from electronic health records
- Developed a distributed tensor decomposition framework that not only scales efficiently with large datasets but also significantly improves the interpretability and robustness of health data analysis, setting a new standard in healthcare analytics.

○ **Precision Medicine via ML-based Causal Inference**

- Spearheaded the development of cutting-edge machine learning methodologies aimed at revolutionizing precision medicine. This involved the intricate use of advanced analytical techniques to tailor patient care based on longitudinal observational data.
- Innovated in the realm of causal sparse tensor estimation, deriving novel theoretical results that provide deeper insights into the complex relationships within high-dimensional and sparse healthcare data.
- Developed a novel generative model specifically for Electronic Health Record (EHR) digital twins simulation, paving the way for more accurate and personalized patient care models and simulations in medical research.

---

## Teaching Experience and Service

2023 **Organizer of ICLR Workshop, Topic: Generative AI and Foundation Models in Health**, Proposal Submitted.

2023 **Co-supervisor to two undergraduate students, Topic: Generative Models for EHRs**, Emory University, Under Review.

2023 Summer **Organizer and Instructor of NSF REU/RET Computational Mathematics for Data Science**, Topic: *Fair Generative Modeling in Healthcare*, Emory University. The work is accepted by 2023 AI In Health Conference

2022 **Organizer of SIAM Conference in Optimization**, Topic: *Optimization for Healthcare*, Seattle.

2022 **Organizer of SIAM Conference on Computational Science and Engineering**, Topic: *Acceleration methods for scientific and machine learning applications*, Amsterdam.

2022 **Technology Chair of Conference on Health, Inference, and Learning (CHIL)**, Cambridge, Massachusetts.

2021 Summer **Instructor of NSF REU/RET Computational Mathematics for Data Science**, Topic: *Learning From Images via Convex Normalizing Flow*, Emory University, Atlanta.

2019-Present **Program Committee/Reviewer of NeurIPS, ICLR, AISTATS, AAAI, AMIA, KDD, TKDE, CHIL, ECCB.**

2019 Spring **Teaching Assistant of Natural Language Processing**, Emory University, Atlanta.

2018, 2019 Fall **Teaching Assistant of Machine Learning**, Emory University, Atlanta.

---

## Awards and Grants

2023 **NSF Travel Grant**, *Workshop on Sparse Tensor Computations*, University of Illinois Urbana-Champaign .

2023 **NSF Travel Grant**, *The fourth Mathematical and Scientific Machine Learning conference*, Brown University .

2023 **Acceleration and Extrapolation methods for Machine Learning**, *SIAM Post-doctoral Support Program*, Under Review .

2012, 2013 **Merit Student Award**, *Shanghai Finance University*.

2011, 2012, 2013, 2014 **First-class scholarship**, *Shanghai Finance University*.

---

## Invited Talks

- 2023 Scientific Computing Meets Machine Learning (Invited Talk). In: *Clemson University*.
- 2023 Multi-Modal EHR Generative model (Invited Talk). In: *INFORMS Annual Data Mining*.
- 2023 Nonlinear Acceleration for minimax and nonlinear optimization (Invited Poster). In: *Brown University*.
- 2023 Optimization for Healthcare. In: *SIAM Conference in Optimization*.
- 2023 Acceleration methods for machine learning applications. In: *SIAM Conference on Computational Science and Engineering*.
- 2022 Acceleration Algorithms for Machine Learning Models (Invited Talk). In: *Weill Cornell Medicine*.
- 2021 Nonlinear Acceleration for Tensor Decomposition. In: *SIAM Conference on Computational Science and Engineering*.
- 2019 Tensor Decomposition for healthcare analytics. In: *2019 Georgia Scientific Computing Symposium*.

---

## Work Experience

- 2021 Summer **Efficient Ensembling Uncertain Quantification**, *Machine Learning Internship*, Lawrence Livermore National Laboratory, Livermore, CA.
  - Aimed to compress deep learning models for better uncertainty quantification.
  - Used iterative pruning methods to train model efficiently.
  - Adopted snapshot learning rate scheme to obtain diverse ensemble models.
  - Adopted focal loss and meta-learning to further enhance uncertainty quantification.
- 01/2021-04/2021 **Continuous Glucose Monitors (CGM) Forecasting**, *Machine Learning Researcher*, SiBionics MedTech, Shenzhen, China.
  - Calibrated model parameters of Continuous Glucose Monitors (CGM) system.
  - Developed a meal detection algorithm using real CGM data.
  - Benchmarked and developed glucose prediction algorithms using real CGM data.