

# BOJIAN HOU

(also Bo-Jian Hou) Address: 4036 Walnut St, Philadelphia, PA, 19104  
Tel: (+1)814-826-8040 ◊ Email: hobo.hbj@gmail.com; bojianh@upenn.edu

## EDUCATION

<b>Nanjing University (NJU)</b> <i>Department of Computer Science and Technology</i> Supervisor: Prof. Zhi-Hua Zhou	<b>Ph.D.</b> <i>Sep.2014-Jun.2020</i>
<b>Nanjing University (NJU)</b> <i>Department of Computer Science and Technology</i>	<b>Bachelor of Science</b> <i>Sep.2010-Jun.2014</i>

## WORK EXPERIENCE

<b>University of Pennsylvania</b> <i>Jan.2022-Present</i>	<b>Postdoctoral Researcher</b>
<b>Cornell University</b> <i>May.2021-Jan.2022</i>	<b>Postdoctoral Associate</b>
<b>4th Paradigm Co.</b> <i>Aug.2020-Apr.2021</i>	<b>Research Scientist</b>

## RESEARCH INTEREST

My research interests focus on trustworthy (generative) artificial intelligence ((Gen)AI) and machine learning (ML), optimization for AI/ML and AI/ML for Healthcare. Specifically, they include:

- **Large Language Models:** developing and applying large language models (LLMs).
- **Multi-model Learning:** developing comprehensive models from multiple modalities.
- **Fairness Learning:** developing fair and unbiased machine learning algorithms.
- **Interpretability:** studying the interpretability of black-box deep learning models.
- **Robust Learning:** developing robust methods to handle open and dynamic environments.
- **Online Learning:** learning models from a continuous stream of data.
- **Biomedical Informatics:** developing machine learning methods to analyze biomedical data (especially for Alzheimer's disease and related dementia (ADRD), and mental health).

## AWARDS & HONORS

<b>Distinguished Paper Award in AMIA 2024 Annual Symposium</b>	2024
<b>PennAITech Innovation Fellow Award</b>	2024
<b>Best Paper Award in the 14th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics</b>	2023
<b>Excellent Doctoral Dissertation Award of Jiangsu Province</b>	2021
<b>Excellent Doctoral Dissertation Award of Nanjing University</b>	2021
<b>JSAI Excellent Doctoral Dissertation Award</b>	2020
<b>CS Excellent Doctoral Dissertation Award of Nanjing University</b>	2020
<b>Outstanding Graduate Student Award of Nanjing University</b>	2020
<b>CCFAI Outstanding Student Paper Award.</b>	2019
<b>National Scholarship for Ph.D.</b>	2017
<b>NeurIPS Volunteer Award</b>	2017

IJCAI Travel Award	2017
Outstanding Undergraduate Student Award of Nanjing University	2014
National Endeavor Scholarship	2012

## PUBLICATIONS ([GOOGLE SCHOLAR](#), 435+ CITATIONS)

Note: \* indicates an equal contribution, and † means the author is my mentee.

**Summary:** 40 full-length peer-reviewed papers are published where 17 of them are first/co-first authored papers. 8 full-length papers are in submission or preparation where 6 of them are first/co-first authored papers. 5 peer-reviewed abstracts are published where 3 are first/co-first authored.

### Full-Length Peer-Reviewed Paper

1. [\[NeurIPS'17\]](#) **Bo-Jian Hou**, Lijun Zhang, and Zhi-Hua Zhou. *Learning with Feature Evolvable Streams*. In: Advances in Neural Information Processing Systems 30, Long Beach, CA, 2017, 30: 1417-1427.
2. [\[IJCAI'17\]](#) **Bo-Jian Hou**, Lijun Zhang, and Zhi-Hua Zhou. *Storage Fit Learning with Unlabeled Data*. In: Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017, 1844-1850.
3. [\[TKDE'19\]](#) **Bo-Jian Hou**, Lijun Zhang, and Zhi-Hua Zhou. *Learning with Feature Evolvable Streams*. IEEE Transactions on Knowledge and Data Engineering, 2019, 33(6): 2602-2615.
4. [\[AAAI'21\]](#) **Bo-Jian Hou**, Yu-Hu Yan<sup>†</sup>, Peng Zhao and Zhi-Hua Zhou. *Storage Fit Learning with Feature Evolvable Streams*. In: Proceedings of the AAAI Conference on Artificial Intelligence, Virtual Conference, 2021, 35(9), 7729-7736.
5. [\[TNNLS'20\]](#) **Bo-Jian Hou** and Zhi-Hua Zhou. *Learning with Interpretable Structure from Gated RNN*. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(7): 2267-2279.
6. [\[TNNLS'21\]](#) **Bo-Jian Hou**, Lijun Zhang, and Zhi-Hua Zhou. *Prediction with Unpredictable Feature Evolution*. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(10): 5706-5715.
7. [\[NeurIPS'23\]](#) Zhuoping Zhou<sup>\*†</sup>, Davoud A. Tarzanagh<sup>\*</sup>, **Bojian Hou**<sup>\*</sup>, Boning Tong, Jia Xu, Yanbo Feng, Qi Long, and Li Shen. *Fair Canonical Correlation Analysis*. In: Proceedings of the 37th Annual Conference on Neural Information Processing Systems, New Orleans, LA, 2023, 36.
8. [\[NeurIPS'24\]](#) Zhuoping Zhou<sup>\*†</sup>, Davoud A. Tarzanagh<sup>\*</sup>, **Bojian Hou**<sup>\*</sup>, Qi Long, and Li Shen. *Fairness-Aware Estimation of Graphical Models*. In: Proceedings of the 38th Annual Conference on Neural Information Processing Systems, Vancouver, Canada, 2024, in press.
9. [\[ISBI'23\]](#) **Bojian Hou**, Hongming Li, Zhicheng Jiao, Zhen Zhou, Hao Zhang, and Yong Fan. *Deep Clustering Survival Machines with Interpretable Expert Distributions*. In: Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI'23), Cartagena de Indias, Colombia, 2023, 1-4.
10. [\[MedIA'24\]](#) **Bojian Hou**<sup>\*</sup>, Zixuan Wen<sup>\*†</sup>, Jingxuan Bao, Richard Zhang, Boning Tong, Shu Yang, Junhao Wen, Yuhua Cui, Jason H Moore, Andrew J. Saykin, Heng Huang, Paul M. Thompson, Marylyn D. Ritchie, Christos Davatzikos, and Li Shen. *Interpretable Deep Clustering Survival Machines for Alzheimer's Disease Subtypes Discovery*. Medical Image Analysis, 2024, 97:103231.
11. [\[UAI'23\]](#) Davoud A. Tarzanagh<sup>\*</sup>, **Bojian Hou**<sup>\*</sup>, Boning Tong<sup>\*†</sup>, Qi Long, and Li Shen. *Fairness-Aware Class Imbalanced Learning on Multiple Subgroups*. In: Proceedings of the 39th Conference on Uncertainty in Artificial Intelligence, Pittsburgh, PA, 2023, 2123-2133.

12. [AMIA IS'24] **Bojian Hou**, Andrés Mondragón<sup>†</sup>, Davoud A. Tarzanagh, Zhuoping Zhou, Andrew J Saykin, Jason H Moore, Marylyn D Ritchie, Qi Long, and Li Shen. *PFERM: A Fair Empirical Risk Minimization Approach with Prior Knowledge*. In: Proceedings of AMIA 2024 Informatics Summit, Boston, MA, 2024, 211.
13. [AMIA IS'24] Weiqing He<sup>\*†</sup>, **Bojian Hou**<sup>\*</sup>, George Demiris, and Li Shen. *Interpretability Study for Long Interview Transcripts from Behavior Intervention Sessions for Family Caregivers of Dementia Patients*. In: Proceedings of AMIA 2024 Informatics Summit, Boston, MA, 2024, 201.
14. [AMIA'24] Jia Xu<sup>\*†</sup>, Tianyi Wei<sup>\*†</sup>, **Bojian Hou**<sup>\*</sup>, Patryk Orzechowski, Shu Yang, George Demiris, Li Shen. *MentalGPT: Harnessing AI for Compassionate Mental Health Support*. In: Proceedings of AMIA 2024 Annual Symposium, San Francisco, LA, 2024, in press.
15. [AMIA'24] Yanbo Feng<sup>\*†</sup>, **Bojian Hou**<sup>\*</sup>, Ari Klein, Karen O'Connor, Jiong Chen, Andrés Mondragón, Shu Yang, Graciela Gonzalez-Hernandez, Li Shen. *Analyzing Dementia Caregivers' Experiences on Twitter: A Term-Weighted Topic Modeling Approach*. In: Proceedings of AMIA 2024 Annual Symposium, San Francisco, LA, 2024, in press.
16. [BigData'24] Weiqing He<sup>\*†</sup>, **Bojian Hou**<sup>\*</sup>, Tianqi Shang<sup>\*</sup>, Davoud A. Tarzanagh, Qi Long, Li Shen. *SEFD: Semantic-Enhanced Framework for Detecting LLM-Generated Text*. In 2024 IEEE International Conference on Big Data, 2024, in press.
17. [CCFAI'19] **Bo-Jian Hou** and Yuan Jiang. *Learning Interpretability from RNN with Feature Evolving*. In: CCF Conference on Artificial Intelligence, Xuzhou, China, 2019. (This paper won the Outstanding Student Paper Award.)
18. [AISTATS'24] Davoud A. Tarzanagh, Parvin Nazari, **Bojian Hou**, Li Shen, and Laura Balzano. *Online Bilevel Optimization: Regret Analysis of Online Alternating Gradient Methods*. In: Proceedings of the 27th International Conference on Artificial Intelligence and Statistics, Palacio de Congresos de València, València SPAIN, 2024, 2854-2862.
19. [ICDM'21] Yi He, Jiaxian Dong, **Bo-Jian Hou**, Yu Wang, and Fei Wang. *Online Learning in Variable Feature Spaces with Mixed Data*. In: Proceedings of the 21st IEEE International Conference on Data Mining, Auckland, New Zealand, 2021, 181-190.
20. [ACMMM'22] Heng Lian, John S. Atwood, **Bo-Jian Hou**, Jian Wu, and Yi He. *Online Deep Learning from Doubly-Streaming Data*. In: Proceedings of the 30th ACM International Conference on Multimedia, Lisbon, Portugal, 2022, 3185-3194.
21. [EMNLP'24] Dawei Li, Shu Yang, Zhen Tan, Jae Young Baik, Sukwon Yun, Joseph Lee, Aaron Chacko, **Bojian Hou**, Duy Duong-Tran, Ying Ding, Huan Liu, Li Shen, Tianlong Chen. *DALK: Dynamic Co-Augmentation of LLMs and KG to answer Alzheimer's Disease Questions with Scientific Literature*. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing, Miami, FL, 2024, in press.
22. [TKDE'23] Heng Lian, Di Wu, **Bo-Jian Hou**, Jian Wu, and Yi He. *Online Learning from Evolving Feature Spaces with Deep Variational Models*. IEEE Transactions on Knowledge and Data Engineering, 2023, 36(8): 4144-4162.
23. [TNNLS'24] Dayang Wang, Feng-Lei Fan, **Bo-Jian Hou**, Hao Zhang, Rongjie Lai, Hengyong Yu, and Fei Wang. *Manifoldron: Direct Space Partition via Manifold Discovery*. IEEE Transactions on Neural Networks and Learning Systems, in press, 2024.
24. [TAI'24] Jing-Xiao Liao, **Bo-Jian Hou**, Hang-Cheng Dong, Hao Zhang, Jinwei Sun, Shiping Zhang, and Feng-Lei Fan. *Quadratic Neuron-empowered Heterogeneous Autoencoder for Unsupervised Anomaly Detection*. IEEE Transactions on Artificial Intelligence, in press, 2024.

25. [JCRD'19] Jie Ren<sup>†</sup>, **Bojian Hou**, and Yuan Jiang. *Deep Forest for Multiple Instance Learning*. Journal of Computer Research and Development, 2019, 56(8): 1670-1676.
26. [BCB'23] Zhuoping Zhou<sup>†</sup>, Boning Tong<sup>†</sup>, Davoud A. Tarzanagh, **Bojian Hou**, Andrew J. Saykin, Qi Long, and Li Shen. *Multi-Group Tensor Canonical Correlation Analysis*. In: Proceedings of the 14th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics, Houston, TX, 2023, 1-10. (This paper won the Best Paper Award.)
27. [AMIA IS'23] Mingquan Lin, Yuyun Xiao, **Bojian Hou**, Tingyi Wanyan, Mohit Manoj Sharma, Zhangyang Wang, Fei Wang, Sarah Van Tassel, and Yifan Peng. *Evaluate Underdiagnosis and Overdiagnosis Bias of Deep Learning Model on Primary Open-Angle Glaucoma Diagnosis in Under-Served Populations*. In: Proceedings of the AMIA 2023 Informatics Summit, Seattle, WA, 2023, 370.
28. [MLMI'23] Boning Tong<sup>†</sup>, Zhuoping Zhou<sup>†</sup>, Davoud A. Tarzanagh, **Bojian Hou**, Andrew J. Saykin, Jason Moore, Marylyn Ritchie, and Li Shen. *Class-Balanced Deep Learning with Adaptive Vector Scaling Loss for Dementia Stage Detection*. In: Proceedings of the 14th International Workshop on Machine Learning in Medical Imaging, Vancouver, Canada, 2023, 144-154.
29. [AMIA IS'24] Ruiming Wu<sup>†</sup>, Bing He, **Bojian Hou**, Andrew J Saykin, Jingwen Yan, and Li Shen. *Cluster Analysis of Cortical Amyloid Burden for Identifying Imaging-driven Subtypes in Mild Cognitive Impairment*. In: Proceedings of AMIA 2024 Informatics Summit, Boston, MA, 2024, 439.
30. [AMIA'24] Boning Tong<sup>†</sup>, Travysse Edwards<sup>†</sup>, Shu Yang, **Bojian Hou**, Davoud A. Tarzanagh, Ryan J. Urbanowicz, Jason H. Moore, Marylyn D. Ritchie, Christos Davatzikos, Li Shen. *Ensuring Fairness in Detecting Mild Cognitive Impairment with MRI*. In: Proceedings of AMIA 2024 Annual Symposium, San Francisco, LA, 2024, in press. (This paper won the AMIA Distinguished Paper Award.)
31. [PSB'25] Kazi Noshin<sup>\*†</sup>, Mary Regina Boland<sup>\*</sup>, **Bojian Hou**, Victoria Lu, Carol Manning, Li Shen, Aidong Zhang. *Uncovering Important Diagnostic Features for Alzheimer's, Parkinson's and Other Dementias Using Interpretable Association Mining Methods*. In: Proceedings of Pacific Symposium on Biocomputing 2025, The Big Island of Hawaii, Hawaii, 2025, in press.
32. [AMIA IS'25] Joseph Lee, Shu Yang, Jae Young Baik, Xiaoxi Liu, Zhen Tan, Dawei Li, Zixuan Wen, **Bojian Hou**, Duy Duong-Tran, Tianlong Chen, Li Shen. *Knowledge-Driven Feature Selection and Engineering for Genotype Data with Large Language Models*. In: Proceedings of AMIA 2025 Informatics Summit, Pittsburgh, PA, 2025, in press.
33. [AMIA IS'25] Tianqi Shang, Shu Yang, Weiqing He, Tianhua Zhai, Dawei Li, **Bojian Hou**, Tianlong Chen, Jason H. Moore, Marylyn D. Ritchie, and Li Shen. *Leveraging Social Determinants of Health in Alzheimer's Research Using LLM-Augmented Literature Mining and Knowledge Graphs*. In: Proceedings of AMIA 2025 Informatics Summit, Pittsburgh, PA, 2025, in press.
34. [AMIA IS'25] Kazi Noshin<sup>\*</sup>, Mary Regina Boland<sup>\*</sup>, **Bojian Hou**, Weiqing He, Victoria Lu, Carol Manning, Li Shen, and Aidong Zhang. *Understanding the Clinical Modalities Important in NeuroDegenerative Disorders, Alzheimer's Disease, and Risk of Patient Injury Using Machine Learning and Survival Analysis*. In: Proceedings of AMIA 2025 Informatics Summit, Pittsburgh, PA, 2025, in press.
35. [AAAI Bridge'25] Kazi Noshin<sup>\*</sup>, Mary Regina Boland<sup>\*</sup>, **Bojian Hou**, Weiqing He, Victoria Lu, Li Shen, and Aidong Zhang. *Integrating Social Determinants of Health in a Multi-Modal Deep Clustering Survival Model for Injury-Risk in Alzheimer's and Related Dementia Patients*. In: Proceedings of AAAI Bridge, Philadelphia, PA, 2025, in press.
36. [AAAI Bridge'25] Bach Nguyen, Tianlong Chen, Shu Yang, **Bojian Hou**, Li Shen, and Duy

- Duong-Tran. *Accessing the Topological Properties of Human Brain Functional Sub-Circuits in Echo State Networks*. In: Proceedings of AAAI Bridge, Philadelphia, PA, 2025, in press. (**This paper won the Best Paper Runners-Up award.**)
37. [SR’22] Mingquan Lin, **Bojian Hou**, Lei Liu, Mae Gordon, Michael Kass, Fei Wang, Sarah H. Van Tassel, and Yifan Peng. *Automated diagnosing primary open-angle glaucoma from fundus image by simulating human’s grading with deep learning*. Scientific Reports, 2022, 12(1): 14080.
  38. [CBM’23] Mingquan Lin, **Bojian Hou**, Swati Mishra, Tianyuan Yao, Yuankai Huo, Qian Yang, Fei Wang, George Shih, and Yifan Peng. *Enhancing thoracic disease detection using chest X-rays from PubMed Central Open Access*. Computers in Biology and Medicine, 2023: 106962.
  39. [iScience’24] Zexuan Wang, Qipeng Zhan, Boning Tong, Shu Yang, **Bojian Hou**, Heng Huang, Andrew J. Saykin, Paul M. Thompson, Christos Davatzikos, and Li Shen. *Distance-weighted Sinkhorn loss for Alzheimer’s disease classification*. iScience, 2024, 27(3): 109212.
  40. [TCBB’24] Zhuoping Zhou<sup>†</sup>, Boning Tong<sup>†</sup>, Davoud A. Tarzanagh, **Bojian Hou**, Andrew J. Saykin, Qi Long, and Li Shen. *MG-TCCA: Tensor Canonical Correlation Analysis across Multiple Groups*. IEEE/ACM Transactions on Computational Biology and Bioinformatics, in press, 2024.

### Full-Length Manuscript

41. [arXiv’21] **Bojian Hou**, Hao Zhang, Gur Ladizhinsky, Ali Kayyal, Stephen Yang, Volodymyr Kuleshov, Fei Wang, and Qian Yang. *Clinical Evidence Engine: Proof-of-Concept For a Clinical-Domain-Agnostic Decision Support Infrastructure*. arXiv:2111.00621, 2021.
42. [Nature MI’25] **Bojian Hou\***, Zhanliang Wang<sup>\*†</sup>, Zhuoping Zhou, Boning Tong, Zexuan Wang, Qi Long, Li Shen. *Fair CCA for Fair Representation Learning: An ADNI Study*. Submitted to Nature Machine Intelligence, 2024.
43. [Innovate Aging’24] **Bojian Hou\***, Weiqing He\*, Amy Zheng, Yanbo Feng, Ari Klein, Karen O’Connor, Shu Yang, Tianqi Shang, Goerge Demiris, Graciela Gonzalez-Hernandez, Li Shen. *Advanced Topic Modeling with Large Language Models: Analyzing Social Media Content from Dementia Caregivers*. Submitted to Innovation in Aging, 2024.
44. [ICLR’25] Ruochen Jin<sup>\*†</sup>, **Bojian Hou\***, Jiancong Xiao\*, Weijie Su, Li Shen. *Fine-Tuning Linear Layers Only Is a Simple yet Effective Way for Task Arithmetic*. arXiv:2407.07089, submitted to ICLR’25, 2024.
45. [TNNLS’25] Ruochen Jin<sup>\*†</sup>, **Bojian Hou\***, Boning Tong, Shu Yang, Li Shen. *ICAFS: Inter-Client-Aware Feature Selection for Vertical Federated Learning*. Submitted to TNNLS, 2025.
46. [ICLR’25] Jia Xu<sup>\*†</sup>, Tianyi Wei<sup>\*†</sup>, **Bojian Hou\***, Patryk Orzechowski, Shu Yang, George Demiris, Li Shen. *MentalChat16K: A Benchmark Dataset for Conversational Mental Health Assistance*. Submitted to ICLR’25, 2025.
47. [BCB’24] Boning Tong<sup>†</sup>, Zhuoping Zhou<sup>†</sup>, **Bojian Hou**, Qi Long, Li Shen. *Balanced Cluster Canonical Correlation Analysis for Alzheimer’s Disease Analysis*. Submitted to ACMBCB, 2024.
48. [arXiv’23] Yan Ma<sup>†</sup>, Weicong Liang, Yiduo Hao, **Bojian Hou**, Xiangyu Yue, Chao Zhang, and Yuhui Yuan. *Revisiting DETR Pre-training for Object Detection*. arXiv:2308.01300, 2023.

### Peer-Reviewed Abstract

49. [AAIC’24] Jia Xu<sup>\*†</sup>, Tianyi Wei<sup>\*†</sup>, **Bojian Hou\***, George Demiris, Li Shen. *Revolutionizing Dementia Care: Enhancing Talk Therapy with Fine-Tuned Large Language Models Using GPT Self-Generated Data*. In Alzheimer’s Association International Conference on Alzheimer’s Disease, Poster 93496. Philadelphia and Online, 2024.



50. [AAIC'24] **Bojian Hou\***, Zixuan Wen<sup>\*†</sup>, Jingxuan Bao, Richard Zhang, Boning Tong, Shu Yang, Junhao Wen, Yuhao Cui, Jason H Moore, Andrew J. Saykin, Heng Huang, Paul M. Thompson, Marylyn D. Ritchie, Christos Davatzikos, Li Shen. *Interpretable Deep Clustering Survival Machines for Alzheimer's Disease Risk Groups Discovery*. In Alzheimer's Association International Conference on Alzheimer's Disease, Poster 92319. Philadelphia and Online, 2024.
51. [AAIC'24] Shu Yang, Austin Wang, Jingxuan Bao, Shizhuo Mu, Yanbo Feng, Zixuan Wen, Jae Young Baik, Junhao Wen, **Bojian Hou**, Rongguang Wang, Heng Huang, Andrew J. Saykin, Paul M. Thompson, Christos Davatzikos, Li Shen. *'Diet' deep canonical correlation analysis for high-dimensional genetics study of brain imaging phenotypes in Alzheimer's disease*. In Alzheimer's Association International Conference on Alzheimer's Disease, Poster 92330. Philadelphia and Online, 2024.
52. [AAIC'24] Yanbo Feng<sup>\*†</sup>, **Bojian Hou\***, Ari Klein, Karen O'Connor, Jiong Chen, Andrés Mondragón, Shu Yang, Graciela Gonzalez-Hernandez, Li Shen. *Exploring Semantic Topics in Dementia Caregiver Tweets*. In Alzheimer's Association International Conference on Alzheimer's Disease, Lightning Presentation Round (LRP) Poster 93035. Philadelphia and Online, 2024.
53. [ICIBM'24] Joanna Ye<sup>†</sup>, Jasmine Zhang<sup>†</sup>, **Bojian Hou**, Li Shen. *Transformer-based Fall Detection Method: A UMAFall Study*. In International Conference on Intelligent Biology and Medicine, 2024.

## PRESENTATIONS & TALKS

---

### Conference Presentation

1. *MentalGPT: Harnessing AI for Compassionate Mental Health Support*, AMIA 2024 Annual Symposium, San Francisco, LA, 2024.
2. *Analyzing Dementia Caregivers' Experiences on Twitter: A Term-Weighted Topic Modeling Approach*, AMIA 2024 Annual Symposium, San Francisco, LA, 2024.
3. *Interpretability Study for Long Interview Transcripts from Behavior Intervention Sessions for Family Caregivers of Dementia Patients*, AMIA 2024 Informatics Summit, Boston, MA, 2024.
4. *PFERM: A Fair Empirical Risk Minimization Approach with Prior Knowledge*, AMIA 2024 Informatics Summit, Boston, MA, 2024.
5. *Cluster Analysis of Cortical Amyloid Burden for Identifying Imaging-driven Subtypes in Mild Cognitive Impairment*, AMIA 2024 Informatics Summit, Boston, MA, 2024.
6. *Fair Canonical Correlation Analysis*. 37th Annual Conference on Neural Information Processing Systems (NeurIPS'23), New Orleans, LA, 2023.
7. *Fairness-Aware Class Imbalanced Learning on Multiple Subgroups*. 39th Conference on Uncertainty in Artificial Intelligence (UAI'23), Pittsburgh, PA, 2023.
8. *Deep Clustering Survival Machines with Interpretable Expert Distributions*, IEEE International Symposium on Biomedical Imaging (ISBI'23), Cartagena de Indias, Colombia, 2023.
9. *Storage Fit Learning with Feature Evolvable Streams*, AAAI Conference on Artificial Intelligence (AAAI'21), Virtual Talk, 2021.
10. *Learning Interpretability from RNN with Feature Evolving*, CCF Conference on Artificial Intelligence (CCFAI'19), Xuzhou, China, 2019
11. *Learning with Feature Evolvable Streams*, Advances in Neural Information Processing Systems 30 (NIPS'17), Long Beach, CA, 2017.

12. *Storage Fit Learning with Unlabeled Data*, 26th International Joint Conference on Artificial Intelligence (IJCAI'17), Melbourne, Australia, 2017.

### Invited Talk

1. *Interpretable Deep Clustering Survival Machines for Alzheimer's Disease Risk Groups Discovery*, ADSP Program Review Meeting, Oct. 2024.
2. *Introduction to MentalGPT*, Lunch Meeting at Naik Lab, University of Pennsylvania, 2024.
3. *Introduction to BERT, GPT, and Finetuning LLM*, BMIN 5210 Class Lecture, University of Pennsylvania, 2024.
4. *Introduction to Attention Mechanism and Transformer*, BMIN 5210 Class Lecture, University of Pennsylvania, 2024.
5. *Fairness-Aware Class Imbalanced Learning on Multiple Subgroups*, BMIN 5210 Class Lecture, University of Pennsylvania, 2024.
6. *Understanding Social Reasoning in Language Models with Language Models*, Lunch Meeting at Shen Lab, University of Pennsylvania, 2024.
7. *Interpretable Deep Clustering Survival Machines for Alzheimer's Disease Risk Groups Discovery*, ADSP Program Review Meeting, March 2024.
8. *Fine-tuning Large Language Models*, Monthly Joint Machine Learning Meeting between CBICA and Shen Lab, University of Pennsylvania, 2023.
9. *Fairness-Aware Class Imbalanced Learning on Multiple Subgroups*, Monthly Joint Machine Learning Meeting between CBICA and Shen Lab, University of Pennsylvania, 2023.
10. *Fairness-Aware Class Imbalanced Learning on Multiple Subgroups*, BMIN 5220 Class Lecture, University of Pennsylvania, 2023.
11. *Machine Learning Methods in Environments with Changing Features*, University of Pennsylvania, 2022.
12. *Machine Learning Methods in Environments with Changing Features*, Cornell University, 2021.
13. *Learning with Interpretable Structure from Gated RNN*, 4th Paradigm Co., 2020.

### ADVISING & MENTORING

---

1. *Yu-Hu Yan (PhD Student, Computer Science, Nanjing University)*. Advised on feature evolvable streaming learning; published in **AAAI'21** [4].
2. *Boning Tong (Female, PhD Student, Bioengineering, UPenn)*. Guided on fairness and imbalance learning; publications in **UAI'23** [11], **AMIA'24 Annual Symposium** [30], and submission to **ACMBCB'24** [47].
3. *Zhuoping Zhou (Female, PhD Student, Applied Mathematics & Computational Science, UPenn)*. Mentored on fairness models for CCA and Graphical Models; published in **NeurIPS'23** [9] and **NeurIPS'24** [10].
4. *Weiqing He (PhD Student, Applied Mathematics & Computational Science, UPenn)*. Advised on interpretability for behavioral interventions published in **AMIA Informatics Summit'24** [13] and AI text detection published in **IEEE Big Data 2024** [16].
5. *Zixuan Wen (Female, PhD Student, Applied Mathematics & Computational Science, UPenn)*. Guided on Alzheimer's disease subtype discovery and survival analysis; published in **Medical Image Analysis** [8].

6. *Jia Xu (Female, Master's Student, Data Analytics, UPenn) & Tianyi Wei (Female, Master's Student, Data Science, UPenn)*. Advised on MentalGPT for mental health; accepted at **AMIA'24 Annual Symposium** [14], **AAIC'24** [49], with submission to **ICLR'25** [46].
7. *Andrés Mondragón (Hispanic American, Undergraduate Student, Economics, Data Science & Statistics, UPenn)*. Advised on fairness work; published in **AMIA Informatics Summit'24** [12].
8. *Yanbo Feng (Master's Student, Bioengineering, UPenn)*. Mentored on dementia caregiver analysis; published in **AMIA'24 Annual Symposium** [15].
9. *Ruochen Jin (Master's Student, Software Engineering, ECNU)*. Advised on disentanglement in task arithmetic; submitted to **ICLR'25** [44].
10. *Zhanliang Wang (Master's Student, Applied Mathematics, UPenn)*. Advised on fair CCA [42] and large language model calibration with submission planned to **JASA**.
11. *Qipeng Zhan (PhD Student, Applied Mathematics & Computational Science, UPenn)*. Mentored on vision language models for Alzheimer's diagnostic tool, with submission planned to **Nature Communication**.
12. *Zeruan Wang (PhD Student, Applied Mathematics & Computational Science, UPenn)*. Advised on multi-agent project using Autogen for explainable AI results, with submission planned to **Nature Communication**.
13. *Riya Mathur (Indian American, Female, Master's Student, Bioengineering, UPenn)*. Guided on mental health data analysis using large language models.
14. *Amy Zheng (Chinese American, Female, Undergraduate Student, Economics & Computer Science, UPenn)*. Advised on dementia caregiver tweet analysis using large language model, submitted to **Innovation in Aging** [43].
15. *Jiong Chen (PhD Student, Bioengineering, UPenn)*. Mentored on graphical models for brain fMRI data analysis, with submission planned to **Nature Methods**.
16. *Joanna Ye & Jasmine Zhang (Chinese American, Female, High School Students)*. Advised on Transformer for elderly fall prediction; accepted at **ICIBM'24** [53].
17. *Travyse Edwards (Black American, PhD Student, Genomics & Computational Biology, UPenn)*. Guided fairness learning on ADNI data; published in **AMIA'24 Annual Symposium** [30].
18. *Ruiming Wu (PhD Student, GCB, UPenn)*. Advised on clustering Alzheimer's risk data; published in **AMIA Informatics Summit'24** [29].
19. *Handi Zhang (Female, Research Assistant, Applied Mathematics, UPenn)*. Advised on scalable federated learning, with submission planned to **ICML'25**.
20. *Richard Zhang (Chinese American, Undergraduate Student, Math & CIS, UPenn)*. Advised on Alzheimer's prediction; published in **Medical Image Analysis** [8].

## GRANT EXPERIENCE

---

1. NIH K99/R00 (PA-24-194), "Enhancing Trustworthy AI for Alzheimer's Disease Prediction", under review.

**Name of PI:** Bojian Hou

**Role:** Principle Investigator

**Project narrative:** This project aims to revolutionize Alzheimer's Disease (AD) management by integrating Artificial Intelligence (AI) in diagnosis, prognosis, and treatment, focusing on enhancing fairness, interpretability, and robustness. By developing advanced AI models and systems,



the project seeks to improve patient outcomes, facilitate personalized medicine, and provide equitable, transparent and robust healthcare solutions. The anticipated impact includes setting new standards for AI in healthcare, influencing global health policies, and significantly advancing the quality of care for AD patients.

2. NIH R01, “Scalable, Fair Machine Learning Models for Integration of Imaging-Omics Data in Alzheimer’s Disease”, under review.

**Name of PI:** Li Shen, Qi Long

**Role:** I helped writing one of the aims and providing preliminary results by designing concrete algorithms and conducting extensive experiments.

**Project narrative:** This research proposal aims to enhance early Alzheimer’s disease (AD) detection using multimodal data analysis and canonical correlation analysis (CCA). The project outlines five aims: developing fair and balanced CCA methodologies, creating scalable approaches for large datasets, developing user-friendly software tools, and applying these methods to real-world AD datasets. It addresses key challenges in CCA, including bias across sensitive attributes, label imbalance, and scalability issues. The research incorporates sketching and federated methods to handle large datasets while preserving privacy. By tackling these challenges, the project seeks to improve early AD diagnosis and advance understanding of disease progression, ultimately contributing to better patient outcomes and potential therapeutic interventions.

3. NIH R01, “Improving the Diagnosis of Alzheimer’s Disease and Related Dementias via Collaborative Multi-view LLM Agents”, under reivew.

**Name of PI:** Li Shen, Tianlong Chen, Qi Long, Kyra S. O’Brien, Ying Ding

**Role:** I helped discussing ideas, writing third of the proposals, designing concrete algorithms and conducting experiments for preliminary results.

**Project narrative:** This research project aims to improve Alzheimer’s disease (AD) diagnosis using advanced AI techniques. It proposes developing unimodal AD Large Language Models (LLMs) to capture knowledge from various data types, including biomarkers, genetics, and medical imaging. These models will then be integrated into a multimodal AI framework using mixture-of-experts, multi-agent collaboration, and collaborative communication approaches. The project will validate these models through real-world use cases, focusing on AD diagnosis, prognosis, and reasoning. By leveraging state-of-the-art LLM agents and collaborative frameworks, the research aims to enhance multi-view analysis of AD, potentially leading to more effective diagnosis and personalized treatment strategies.

4. NSF SCH, “Social Determinants of Health Enhanced Machine Learning Models for Early Prediction of Alzheimer’s Disease and Related Dementias”, under review.

**Name of PI:** Aidong Zhang, Carol Manning, Mary Regina Boland, Li Shen, Kishlay Jha, Shana Stites

**Role:** I helped discussing ideas, writing one of the four tasks of the proposals, designing concrete algorithms and conducting experiments for preliminary results.

**Project narrative:** The research project aims to enhance early Alzheimer’s Disease and Related Dementias (ADRD) prediction by incorporating social determinants of health (SDoH) factors. These factors, which account for 30-55% of health morbidity, are often buried in unstructured clinical notes. The project addresses three main challenges: automatically extracting SDoH from electronic health records, designing SDoH-enhanced interpretable machine learning models for ADRD prediction, and inferring associations between SDoH and ADRD biomarkers. This approach seeks to improve prediction accuracy and timeliness, potentially transforming early de-

tection and management of complex neurological disorders while promoting equitable healthcare access.

5. NIH R01, “Machine Learning Algorithms to Address Health Disparities in Alzheimer’s Risk Prediction from Multi-Modal Clinical Data”, under review.

**Name of PI:** Li Shen, Mary Regina Boland, Duy Duong-Tran

**Role:** I helped discussing ideas, writing one of the three aims of the proposals by leveraging the materials from my published paper.

**Project narrative:** Alzheimer’s Disease and Related Dementias (ADRD) is a significant public health crisis marked by disparities in race, ethnicity, sex, and socioeconomic status. We propose innovative fair machine learning algorithms to address these health disparities in AD risk prediction, leveraging multi-modal, community-based clinical data to improve clinical decision support. The resulting methods are expected to broadly impact biomedical research and benefit public health.

## PROFESSIONAL ACTIVITIES

---

### Program Committee (PC) Member or Reviewer of Conferences

- PC Member of ICLR’25 (The Thirteenth International Conference on Learning Representations)
- PC Member of NeurIPS’24 (The 38th Annual Conference on Neural Information Processing Systems)
- PC Member of ICML’24 (The 41th International Conference of Machine Learning)
- PC Member of NeurIPS’23 (The 37th Annual Conference on Neural Information Processing Systems)
- PC Member of NeurIPS’22 (The 36th Annual Conference on Neural Information Processing Systems)
- PC Member of ICML’22 (The 39th International Conference of Machine Learning)
- PC Member of NeurIPS’21 (The 35th Annual Conference on Neural Information Processing Systems)
- SPC Member of IJCAI’21 (The 30th International Joint Conference on Artificial Intelligence)
- PC Member of ICML’21 (The 38th International Conference of Machine Learning)
- PC Member of AAAI’21 (The 35th AAAI Conference on Artificial Intelligence) (**I was selected as one of the top 25% PC members at this conference.**)
- Reviewer of KDD’20 (The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining)
- PC Member of NeurIPS’20 (The 34th Annual Conference on Neural Information Processing Systems)
- PC Member of AISTATS’20 (The 23rd International Conference on Artificial Intelligence and Statistics)
- PC Member of ICLR’20 (The 8th International Conference on Learning Representations)
- PC Member of AAAI’20 (The 34th AAAI Conference on Artificial Intelligence)
- PC Member of DFM’19 (The 1st ICDM Workshop on Dynamic Feature Mining)
- Reviewer of PRICAI’19 (The 16th Pacific Rim International Conference on Artificial Intelligence)
- PC Member of CCML’19 (The 17th China Conference on Machine Learning)
- PC Member of NeurIPS’19 (The 33rd Annual Conference on Neural Information Processing Systems)
- PC Member of ICML’19 (The 36th International Conference of Machine Learning)
- PC Member of AISTATS’19 (The 22nd International Conference on Artificial Intelligence and Statistics)
- PC Member of ICLR’19 (The 7th International Conference on Learning Representations)
- PC Member of AAAI’19 (The 33rd AAAI Conference on Artificial Intelligence)
- PC Member of NeurIPS’18 (The 32nd Annual Conference on Neural Information Processing Systems)

### Reviewer of Journals

- Reviewer of TPAMI (IEEE Transactions on Pattern Analysis and Machine Intelligence)
- Reviewer of TNNLS (IEEE Transactions on Neural Networks and Learning Systems)
- Reviewer of MIA (Medical Image Analysis)

- Reviewer of TII (IEEE Transactions on Industrial Informatics)
- Reviewer of Nature Methods
- Reviewer of Machine Learning
- Reviewer of TKDD (ACM Transactions on Knowledge Discovery from Data)
- Reviewer of Scientific Reports
- Reviewer of KIS (Knowledge and Information Systems)
- Reviewer of FCS (Frontiers of Computer Science)

### Other Services

- Web Chair of MLA'16 (The 14th Chinese Workshop on Machine Learning and Applications)
- Web Chair of MLA'15 (The 13rd Chinese Workshop on Machine Learning and Applications)
- Web Chair of LAMDA Group from Sept. 2014 to Sept. 2017 (Including updating each page of LAMDA website, maintaining mail server, maintaining course FTP etc.)

## TEACHING ASSISTANT

---

### Advanced Methods and Health Applications in ML (for graduate students, Penn) Spring, 2024

- Preparing homework and quiz
- Grading homework and quiz
- Preparing and Grading Final Projects
- Lecture: Fairness-Aware Class Imbalanced Learning on Multiple Subgroups
- Lecture: Introduction to Attention Mechanism and Transformer
- Lecture: Introduction to BERT, GPT, and Finetuning LLM

### Advanced Methods and Health Applications in ML (for graduate students, Penn) Spring, 2023

- Preparing homework and quiz
- Grading homework and quiz
- Lecture: Fairness-Aware Class Imbalanced Learning on Multiple Subgroups

### Computational Thinking (for undergraduate students, Nanjing University) Fall, 2016

- Preparing homework and quiz
- Grading homework and quiz
- Lecture: Problem Decomposition: Breaking Down Complex Problems into Manageable Parts

### Introduction to Java (for undergraduate students, Nanjing University) Fall, 2014

- Preparing homework and quiz
- Grading homework and quiz
- Lecture: Getting Started with Java: Variables, Data Types, and Basic Syntax

## TECHNICAL STRENGTHS

---

<b>Programming:</b>	Python, Java, C/C++, MATLAB, Bash, LaTeX, HTML, CSS
<b>Machine Learning Platforms:</b>	PyTorch, TensorFlow, Keras, Pandas, Scikit-Learn, SciPy
<b>Other Tools:</b>	Word, Powerpoint, Excel, Pycharm, VSCode, Origin, Mendeley

## LEADERSHIP

---

**2014 Class Representative of Department of Computer Science and Technology at Nanjing University** Sep. 2010 - Jun. 2014

**President of Graduate English Club of Nanjing University** Sep. 2014 - Jun. 2016