Chen Song

scott.songch@gmail.com

Dear hiring manager,

I am writing to express my interest in a potential position in your lab. I am a graduate student at Temple University, specializing in computer science, artificial intelligence track with research topic of machine learning in healthcare. I hold a master of engineering, with a major in pattern recognition, from the South China University of Technology. I am currently seeking an opportunity to conduct research.

My research interests focus on machine learning applications on biological based tasks, especially genomics based and imaging based algorithm development. I focus on predictive and generative deep learning algorithms related to healthcare applications. In detail, I developed a diffusion generative model based genotype augmentation algorithm (genoDiffusion) to alleviate the scarcity of minority population and overall genetics privacy issues in TCGA prostate cancer genotype data, 1000 genome genotype data (human leukocyte antigen region). I developed a deep non-negative matrix factorization based semi-supervised learning algorithm to predict Alzheimer's disease from 3 human blood gene expression profiling databases and 14 brain tissue gene expression profiling databases. Additionally, I investigated homomorphic encryption (HE) based privacy-preserving approach on deep learning applications on genomics through IDASH PRIVACY & SECURITY WORKSHOP - secure genome analysis competition (2020-2023). A HE-friendly deep learning network (ReActHE) was designed to predict quantitative phenotype traits from genotype, predict virus traits from sequence data, and predict heart failure from healthcare information of patients. A sparseHE approach was developed to mitigate the privacy risk of the HE-friendly model when predicting quantitative phenotype from yeast genotype and predicting pathological stage from genotype of TCGA prostate cancer patients. Besides, I also investigated topological data analysis based approaches to improve the performance and robustness of generative models.

In future studies, I plan to develop novel deep-learning algorithms focusing on bio-image, and genomics data from the perspective of real-world application. My academic experiences and research align closely with the area of expertise of your lab. In this regard, I sincerely hope that I will have an opportunity to join the lab.

I am good at Python, Java, Matlab language, PyTorch platform, TensorFlow platform, machine learning related packages, high-performance computing, and container techniques (docker, singularity, etc.). I have fundamental knowledge of machine learning and mathematics including linear algebra, statistics, probability, optimization, integral transform, etc. I am familiar with genotype, gene expression, brain MRI/DTI images, and health records from TCGA, ADNI, and GEO databases.

Thank you for reading this letter. I enthusiastically look forward to hearing from you.

Best regards,
Chen Song

CURRICULUM VITAE

PERSONAL INFORMATION

Name: Chen Song
Email: scott.songch@gmail.com

EDUCATION

Temple University, Philadelphia, Pennsylvania, USA   **2021.9 - 2024.12**
Graduate Student
Major: Computer and Information Science   **GPA:** 3.88/4.00

South China University of Technology (Top 1% in China), Guangzhou, China   **2017.9 - 2020.6**
Degree: Master of Engineering
Major: Control Engineering   **GPA:** 87.48/100
Thesis: Identification and Rapid Recognition of Spiral Tip via Deterministic Learning Theory

South China University of Technology(Top 1% in China), Guangzhou, China   **2013.9 - 2017.6**
Degree: Bachelor
Major: Automation   **GPA:** 3.19/4.00

PUBLICATIONS

1. <u>Chen Song</u>, Yuzhou Chen, Huanmei Wu, Xinghua Shi. GenoDiffusion: Conditional Denoising Diffusion Model for Genomic Data Augmentation. The 12th International Conference on Computational Advances in Bio and Medical Sciences, 2023

2. <u>Chen Song</u>, Wenkang Zhan, Yuzhou Chen, Xinghua Shi. Topo-Diffusion: Topological Diffusion Generative Models. In process.

3. <u>Chen Song</u>, Wenkang Zhan, Xinghua Shi. SparseHE: an efficient privacy-preserving biomedical prediction approach using sparse homomorphic encryption. IEEE International Conference on Healthcare Informatics (ICHI) 2024.

4. Junjie Chen, Jiahao Li, <u>Chen Song</u>, Bin Li, Qingcai Chen, Hongchang Gao, Wendy Hui Wang, Zenglin Xu, Xinghua Shi. Discriminative Forests Improve Generative Diversity for Generative Adversarial Networks. The Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24)

5. Wenkang Zhan, <u>Chen Song</u>, Yang Zhao, Yuzhou Chen, Xinghua Shi. SAT-Diffusion: A Self-Adaptive Topological Diffusion Model Improves Generative Tasks. In process

6. Wenkang Zhan, <u>Chen Song</u>, Yang Zhao, Bin Li, Yuzhou Chen, Hongchang Gao, Wendy Hui Wang, Xinghua Shi. Defend against Membership Inference Attack When Topological Data Analysis Meets Generative Models. In process

7. <u>Chen Song</u>, Wenkang Zhan, Yang Zhao, Bin Li, Yuzhou Chen, Xinghua Shi. TIDM: A Topological Inverse Diffusion Model for Image Restoration. In process.

8. Wenkang Zhan, <u>Chen Song</u>, Supratim Das, Timothy R. Rebbeck, and Xinghua Shi. E2EGraph: An End-to-end Graph Learning Model for Interpretable Prediction of Pathological Stages in Prostate Cancer. bioRxiv (2023): 2023-03.

9. <u>Chen Song</u>, Xinghua Shi. ReActHE: A homomorphic encryption friendly deep neural network for privacy-preserving biomedical prediction. Smart Health.

10 <u>Chen Song</u>, Hongchang Gao, Chang Su, Bari J Dzomba, Huanmei Wu, Xinghua Shi. Semi-supervised Deep Non-negative Matrix Factorization of Gene Expression Profiles for Alzheimer's Disease Prediction, In Process.

11 Xunde Dong, <u>Chen Song</u>, Cong Wang. Spiral Tip Identification via Deterministic Learning. International Journal of Bifurcation and Chaos, 2019, 29(3):1950040.

12 <u>Chen Song</u>, Xunde Dong , Cong Wang. Spiral Tip Recognition via Deterministic Learning. International Journal of Bifurcation and Chaos, 2020, 30(6):2050093.

13 <u>Chen Song</u>, Xunde Dong, Cong Wang. A Frequency-Speed Subsystem of the Barkley Model for Spiral Tip Identification via Deterministic Learning, 2019 Chinese Control And Decision Conference (CCDC). 2019.

RESEARCH TOPICS

Trustworthy, privacy-preserving machine learning (ML) on healthcare data including records, genotype, sequence data

- Developed a privacy-preserving homomorphic encryption (HE) based ML algorithm to defend against membership inference attacks (MIA) on the model sparsity by tightening generalization gap on phenotype prediction with genotype and on relative detection in Forensic database [3].

- Developed a deep learning friendly HE model, ReActHE network, for precise secure multi-label tumor classification on genetic variant and for viral strain classification on genome [9].

- Integrated high-level invariant features obtained by topology data analysis (TDA) into generative models, including denoising diffusion probabilistic model (DDPM), generative adversirial network (GAN), etc., to defend against MIA on data augumentation tasks [6].

- Developed a conditional DDPM algorithm to alleviate the data scarcity of genomics from minority populations in databases, and protect genetic privacy on genotype data augumentation [1].

- Developed a graph neural network (GNN) to model co-effects in prostate gene expression for pathological stage classification; employed post-hoc interpretable ML strategies and biological interpretation to explain decision-making of the proposed GNN [8].

- Developed a non-negative matrix factorization based semi-supervised algorithm to predict Alzheimer's disease from gene expression profiles[10].

Deep learning algorithms on data generation and augumentation

- Enhanced image generation by improving likelihood estimation and accelerating convergence via modeling TDA feature into DDPM [2].

- Developed a dynamic self-adaptive network to compensate for TDA estimation error during deep model training [5].

- Developed a TDA stochastic inverse model with pretrained DDPM generative model on imputing missing image [7].

- Developed a forest GAN model to improve generative diversity on data augumentation [4].

Master thesis: **Identification and rapid recognition of spiral tip via deterministic learning theory**

- Used radial basis function (RBF) network to identify the dynamic of spiral tip from Barkley model [11].

- Developed rapid recognition of spiral tip from the dynamic which is stored in RBF weights [12].

- Built a frequency-speed system from partial differential Barkley model of spiral tip for rapid identification [13].

TEACHING EXPERIENCE

Lab of **CIS 1068 Program Design and Abstraction**	**Fall 2023, Fall 2022**

WORKING EXPERIENCE

Research Assistance at Temple University	**2020.7 - 2021.5**

PRESENTATION EXPERIENCE

International Conference on Intelligent Biology and Medicine (Philadelphia, USA)	**2022**
The 9th Annual MidAtlantic Bioinformatics Conference (Philadelphia, USA)	**2022**
The 10th Annual MidAtlantic Bioinformatics Conference (Philadelphia, USA)	**2023**
The 12th International Conference on Computational Advances in Bio and Medical Sciences (Oklahoma, USA)	**2023**

PROGRAM TECHNIQUES

Python; Matlab; Java; C ++; R language; Pytorch; Tensowflow; High Performance Computing; Container (Docker,Singularity); Latex

RELATED COURSES

CIS 5526 Machine Learning	A
CIS 5590 Optimization of Machine Learning	A
CIS 5523 Knowledge Discovery & Data Mining	A-
CIS 5590 Machine Learning for Bioinformatics and Biomedical Data	A
CIS 5525 Neural Computation	A
CIS 5603 Artificial Intelligence	A-
CIS 5543 Computer Vision	A-

REVIEWER EXPERIENCE

The 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining	**2024**
The 13th IEEE International Conference on Healthcare Informatics	**2024**

LANGUAGES

English: proficiency
Chinese: native speaker

AWARDS

Undergraduate:	
Award for Study Progress in University	**2016**
Graduate:	
Third-class Scholarship in University	**2017**
First-class Scholarship in University	**2018**
First-class Scholarship in University	**2019**