

XIAOYAN LI

🏠 2385 Burnt Tree Ln, Apt 5, East Lansing, MI 48823 📞 (412)-418-5816

✉️ lixiaoy5@msu.edu 🌐 <https://www.linkedin.com/in/xiaoyan-li>

🎓 EDUCATION

Michigan State University, GPA: 3.79/4.00	Expected Dec 2022
Ph.D. in Computational Mathematics, Science & Engineering	
MS in Computer Science & Engineering	
Carnegie Mellon University, MS in Civil & Environmental Engineering	Jan 2015
China Agricultural University, BS in Agricultural Engineering	Jun 2013

💼 EXPERIENCE

Intern Software Developer-Lucy Labs, New York 🌐	2022
• Developed quantitative trading algorithms using various mathematical knowledge and programming skills. Utilized various traditional financial markets' quantitative strategies along with development of various trading systems.	
Project leader-MSU, East Lansing 🌐	May 2018 - Present

Water Contaminants Quantification by Transformer-U-Net model project

- Developed a **novel low-cost (from hundred dollars to one dollar) and fast (from days to seconds) deep learning model (Transformer-U-Net CNN)** to **recognize and quantify** tap water contaminant concentrations based on cell phone image. Built the model by **transfer learning** with 625 Coffee-Ring Effect nanochromatography images and corresponding 4200 Energy-dispersive X-ray spectroscopy (EDS) mappings.
- Project awarded \$6700 from CEE department

Intern Data scientist-Xie's AI Lab, Lansing, MI 🌐	May 2021 - Present
---	--------------------

- Developed a **Heterogeneous Graph Neural Network (GNN)** model to remove **batch effects in complex, high dimensionality** single-cell-seq datasets integration. Built the **first model** by incorporating gene pathway in message passing information to improve gene expression prediction.
- Developed **MultiDataset-MultiTask** learning algorithm for building model on multiple datasets decreased protein expression RMSE error from 1.19 to 0.51.

🔬 RESEARCH & PROJECTS

Tap Water Fingerprinting by CNN model with Coffee-Ring Effect 🌐	Jul 2016 - Aug 2018
---	---------------------

Tech Stack: Deep learning, Pytorch, CNN, K-means clustering

- Built a CNN model to classify water samples nanochromatography pattern into 6 classes and achieved comparable accuracy as human clustering on water treatment method with **76.7 ± 3.0%** accuracy.

Comparison of Classical and CNN methods on Human Expression recognition 🌐	Sep 2018 - Dec 2018
---	---------------------

Tech Stack: Pytorch, SVM, Gaussian Process Classifier, Logistic Regression

- Collaborated with teammates to build and tested Logistic Regression Classifier (**60.1%**), Multi-layer Perception Classifier (**60.4%**), SVM Classifier (**60.2%**), CNN Classifier (**90.0%**) on the CK+ (1.7GB) and AffactNet (55GB) dataset with image gray-scale transformation and Gaussian Blur.

Ca ²⁺ video data enhancing through 3D CNN Res-Net model	Jan 2020 - May 2020
--	---------------------

Tech Stack: 3D CNN model implementation, video/image data noise removal

- Designed and implemented the 3D Res-Net CNN model to extract the mouse brain neuron activity features and remove the instrument background noise which achieved 19.387 db.

Database Functions Implementation 🌐	May 2021 - Aug 2021
-------------------------------------	---------------------

Tech Stack: SQLite, Wait-die/Wound-wait scheduler, loss-join, Conflict serializable

- Implemented a validation-locking schedule function to validate the schedule of the legality, two-phased locking, and consistency errors in the actions and the conflict-serializable function to verify the schedule is serializable.
- Implemented a transaction concurrency control scheduler based on wait die protocol.

🏆 AWARDS AND HONORS

NeurIPS 2021 competition: Single-cell Analysis competition (Team name : DANCE) 🌐	Nov 2021
--	----------

- Rank one in Modality prediction task
- Collaborated with colleagues with different backgrounds (computer science, statistics and biology) developed the first graph neural network(GNN) to predict RNA expression based on chromatin accessibility achieved the overall lowest RMSE 0.2835.

🔧 SKILLS

Languages: Python, C, C++ , HTML/CSS, BASH, MATLAB, R

Database: MySQL, SQLite

Framework: Keras, Tensorflow, Pytorch, Numpy, Pandas, CV2, Deep Graph Library, Seurat

Tools & Technologies: AWS S3, Docker, Git, Hadoop, Computer vision, Boosting, XGBoost, multimedia dataanalysis, object-detection, image/video denoising

📄 SELECTED PUBLICATIONS

- Li X, Sanderson AR, Allen SS, Lahr RH. Tap water fingerprinting using a convolutional neural network built from images of the coffee-ring effect. Analyst. 2020; 145(4):1511-1523. doi:10.1039/c9an01624d
- Wang X, Wang W, Lowry G et al Preparation of palladized carbon nanotubes encapsulated iron composites: highly efficient dechlorination for trichloroethylene and low corrosion of nanoiron. R Soc Open Sci 5:172242. <https://doi.org/10.1098/rsos.172242>
- Li X, Su Runze, Xie Yuying. scRNA-seq data batch effect removal by customized Heterogeneous transformer graph neural network. (Manuscript)