Zhouyayan (Iris) Li

Ph.D., Engineering
University of Iowa

Cell: (319) 512-3523
Email: zhouyayan.li@gmail.com
LinkedIn: linkedin.com/in/Zhouyayan-li

Professional Statement

I am an experienced data scientist with a strong background in real-world data problem-solving. I have been working with multi-sourced large-scale structured and unstructured datasets since 2020. I have a proven track record in solving real-world complex environmental and geospatial problems with data science analytical skills and cloud & web application development capability.

Education

Ph.D. in Engineering (Water Resources Management), University of Iowa, USA, May 2024

B.S. in Engineering, Hohai University, China, July 2018

Experience

Data Scientist, Full-time Contract

John Deere, February 2024 – Current

- Developed and maintained data pipelines to process and extract insights from large-scale agricultural, geospatial, and machine data (at the trillion-level) utilizing Azure Databricks, AWS, and Git.
- Developed data visualizations and application products from multi-modal datasets using Python, SQL, and open-source geospatial libraries such as GDAL, Shapely, H3, and Folium, leveraging machine learning, deep learning, and statistical models.
- Led a team of three on two geospatial R&D projects, integrating data from remote sensing and in-field measurements to support decision-making and create customer-facing products.
- Filed a patent based on a geospatial R&D project I led; delivered presentations and demos to both internal and external stakeholders.
- Secured 3rd place in a company-wide innovation competition (555 total submissions) for a project that enhanced the spatial resolution of public remote sensing images using deep models, such as GAN, in the customer satisfaction category.
- Developed TestGenie, a Gen-AI powered tool for manipulating data via natural language commands, building 80% of the core modules; led a team of 5 to enhance UI design, backend-frontend integration, and result export features during a company innovation event.

Skills Involved: deep learning, machine learning, remote sensing, Azure, AWS, Docker, GitHub, big data, leadership, communication, R&D, pipeline management, Gen-AI, LLM.

Data Scientist / Research Assistant

University of Iowa Hydroinformatics Lab, October 2020 – February 2024

- Customized and developed production-ready machine learning & deep learning models using Python and PyTorch framework for environmental and geospatial Computer Vision (CV) and time-series forecasting (NLP) tasks.
- Innovated a CNN-NLP-coupled architecture using deep CNN, MLP, and natural language modeling techniques, such as Transformers, surpassing benchmark performances in Earth surface forecasting by 31.9%.
- Enhanced prediction accuracy by over 5% through a unique sequence and time-series module integrating Transformer, MLP, and Attention techniques. Introduced a CNN-NLP-coupled model to optimize the processing of spatially neighboring data.
- Acquired advanced data modeling techniques such as normalization, data quality inspection, coverage assessment, attribute analysis, and performance management for both classification and regression tasks.
- Executed data mining and feature engineering with statistical probabilistic inference, significantly reducing computational costs (hardware and time) by over 70%.
- Worked with various Key Performance Indicators (KPI), such as F1-Score, Dice score, MAE, RMSE, and NSE. Acquired hands-on experience in defining, quantifying, and analyzing KPIs that define successful customer outcomes.

- Led and mentored teams (4+ members) in experiment design, debugging, and result presentation, fostering collaboration and skill development among junior team members.

Skills Involved: CVML, time series forecasting, natural language models, machine learning, developing production-ready models, Python, analyzing Key Performance Indicators, data modeling techniques, leadership.

Geospatial Data Analyst / Research Assistant
University of Iowa Hydroinformatics Lab, August 2020 – February 2024

- Specialized in remote sensing image processing, working with high resolution Landsat 8, Sentinel 2, PlanetScope, and C-band SAR imagery, alongside 10+ rainfall and soil moisture datasets. Expert in detecting surface anomalies on Earth.
- Engineered 40+ automated data pipelines, utilizing Python, GIS programming, and SQL for TB-scale data-access and manipulation. These pipelines efficiently produce million-record clean dataset samples ready for analysis.
- Developed 10+ interactive geospatial visualizations employing Tableau, D3, and Observable to effectively communicate findings and methodologies to stakeholders with a variety of backgrounds.
- Experienced working with diverse teams. Collaborated as a consultant with NOAA, NASA, and Iowa Department of Transportation on over three significant projects in natural hazard mapping and environmental sustainability.
- Created an accessible, large-scale Earth surface monitoring tool integrating Esri ArcGIS and Cloud platforms, such as Google Earth Engine and GCP, enhancing online comparative analysis and research.

Skills Involved: remote sensing data, GIS, SQL, Python, visualization tools (Tableau and D3 JS), Cloud computing and platforms, excellent communication skills diverse stakeholders with a variety of backgrounds.

Web-Based GIS System Developer Intern
Iowa Flood Center, May 2019 – July 2020

- Accessing, cleaning, preprocessing, and analysis of hazard, risk, and geospatial datasets, both structured and unstructured, in vector and raster formats from distributed sources, ensuring high data accuracy and reliability.
- Developed a web-based natural hazard information application capable of generating flood inundation maps on standard personal computers. Utilized JavaScript, HTML, CSS, and PHP for robust and responsive design.
- Effectively communicated non-technical aspects of the application to diverse audiences in Iowa, offering data solutions and analytical support for Earth surface mapping and natural hazard awareness.
- Engaged in cross-functional collaborations in large-scale geospatial analytics projects, contributing to effective decision-making and rapid response strategies.

Skills Involved: JavaScript, vector and raster data structures, Geospatial data search and analysis, Distributed Datasets, structured and unstructured data, communication to nontechnical audiences.

Additional Leadership Experience

- Spearheaded the organizational committee for the University of Iowa Hackathon for over three years (2021 - 2023), successfully orchestrating a nationwide event that catered to the diverse needs of 700+ participants from various universities and industries across the United States.
- Led a multidisciplinary team to ensure seamless execution of U of I Hackathon, coordinating logistics, participant engagement, and resource allocation. Fostered a collaborative environment that encouraged innovation and creativity among participants.

Certifications

- DevOps on AWS Specialization — July 2024, License HSTFVZWJFG3P
- Deep Learning Specialization — August 2022, License Q3N3FMHYDL5A
- Machine Learning — October 2020, License B464WQTTBHJS
- Google Data Analytics Specialization — June 2022, License UQVNEW6AGQA4

Selected First-Authored Peer-Reviewed Journal Publications

Li, Z., & Demir, I. (2024). Better localized predictions with Out-of-Scope information and Explainable AI: One-Shot SAR backscatter nowcast framework with data from neighboring region. *ISPRS Journal of Photogrammetry and Remote Sensing*, *207*, 92–103. https://doi.org/10.1016/J.ISPRSJPRS.2023.11.021

Li, Z., & Demir, I. (2024). MultiRS flood mapper: a google earth engine application for water extent mapping with multimodal remote sensing and quantile-based postprocessing. *Environmental Modelling & Software*, *176*, 106022. https://doi.org/10.1016/J.ENVSOFT.2024.106022

Li, Z., Xiang, Z., Demiray, B. Z., Sit, M., & Demir, I. (2023). MA-SARNet: A one-shot nowcasting framework for SAR image prediction with physical driving forces. *ISPRS Journal of Photogrammetry and Remote Sensing*, *205*, 176–190. https://doi.org/10.1016/J.ISPRSJPRS.2023.10.002

Li, Z., & Demir, I. (2023). U-net-based semantic classification for flood extent extraction using SAR imagery and GEE platform: A case study for 2019 central US flooding. *Science of the Total Environment*, *869*. https://doi.org/10.1016/j.scitotenv.2023.161757

Li, Z., & Demir, I. (2022). A comprehensive web-based system for flood inundation map generation and comparative analysis based on height above nearest drainage. *Science of The Total Environment*, *828*, 154420. https://doi.org/10.1016/J.SCITOTENV.2022.154420

Skills Involved: Written communication, formal analysis report writing, R&D.

Selected Conference Presentations

"Satellite Imagery Enhancement", John Deere Tech Conference, Des Moines, Sept. 11 – Sept. 12, 2024.

"MA-SARNets: One-Shot SAR Backscatter Nowcast Frameworks with Physical Driving Forces and Out-of-Scope Information", AGU Fall Meeting, San Francisco, Dec. 12 – Dec. 15, 2023.

"Flood Sequence Mapping with Multimodal Remote Sensing under the Influence of Dense Vegetation: Mississippi River Case Study", AGU Fall Meeting, San Francisco, Dec. 12 – Dec. 15, 2023.

"Flood Mapping with Multimodal Remote Sensing Imagery: Application and Adjustment", Iowa Flood Center Meeting, Iowa City, Aug. 30, 2023.

"U-Net-Based Semantic Classification for Flood Extent Extraction Using SAR Images: A Case Study for 2019 Central US Flooding", AGU Fall Meeting, Chicago, Dec. 12 – Dec. 17, 2022.

"Understanding the Mechanisms behind the Performance of HAND, A Simplified Inundation Mapping Technique", Iowa Center Flood Meeting, Iowa City, Feb. 2, 2022.

"A Comprehensive Web-based System for Flood Inundation Mapping and Analysis Based on Height Above Nearest Drainage", AGU Fall Meeting, New Orleans, Dec. 13 – Dec. 17, 2021.

Skills Involved: Communication (oral and written)