Haotian Wang

(860)-234-6647 | LinkedIn | google scholar | haw23012@gmail.com

PROFESSIONAL SUMMARY

Highly motivated PhD students with 5+ years of experience in data science and **machine learning**. Seeking a DS position to leverage my large-scale data processing, experiment design, and ML model development skills.

EDUCATIONAL BACKGROUND

Ph.D. in Engineering, **University of Connecticut** Storrs, CT. 2023 – Dec 2026

*Published **4** papers in prestigious journals and **5** in conferences; Received CUAHSI **Machine learning fellowship ($10,000)***

M.Sc. in Engineering, **Wuhan University** Wuhan, China. 2020 - 2023

B.S. in Engineering, **Wuhan University** Wuhan, China. 2016 – 2020

TECHNICAL CAPABILITIES

- Languages: Python, R, SQL, MATLAB, Scala, C#, C, C++, JavaScript
- ML toolkits/Models: Pytorch, Scikit-learn, Keras, Tensorflow, Hugging Face, NumPy, SpaCy, NLTK, LR, RF, XGBoost, LightGBM, CatBoost, ARIMA, LSTM, GRU, CNN, GCN, RAG, LLM
- Other skills: Airflow/Spark, Databricks, Snowflake, Jupyter, Dash, Streamlit, AWS, GCP, Azure, Apache, Tableau**,** Git, Docker

WORK AND RESEARCH EXPERIENCE

National Water Center (NWC) Tuscaloosa, AL

Data Scientist Intern *Jun 2025 - Aug 2025*

- Applied propensity score matching and **ML models** to identify the key drivers of monitor site optimization. Designed and executed A/B experiments on the DL-based model, resulting in a 10% accuracy improvement and reducing 30% labor costs.
- Diagnosed data sparsity as a key bottleneck in causal inference and implemented a **multi-modal augmentation model (GAN, Vertex AI)**, generating 100k samples through generative deep learning models, delivering 5x efficiency gain.
- Designed and implemented a hybrid ML (**GAN, U-net, LLM**) workflow to segment and detect flood alarms through stratified sampling of catchments, successfully deployed with high generalization ability.
- Developed a **dashboard** using Dash, matplotlib, and plotly for visualization to report to stakeholders weekly, published 2 papers in a top-tier conference. Collaborated with the R&D team, PM team, and data scientist team actively.

University of Connecticut Storrs, CT

Research Assistant *Jun 2023 - Present*

- Diagnosed imbalance-driven errors in flood forecasting and built a two-tower probabilistic machine learning (**RF, XGBoost, LightGBM**) model that lifted Pearson correlation by 16%, documented in a first-author publication in a top-tier journal.
- Built an automated **LSTM** predictive system for 200+ datasets for weather forecasting, achieving 0.902 Pearson R without relying on large-scale data dependency, saving ~100 labor hours each year.
- Built end-to-end preprocessing pipelines in Python to clean and balance 2B+ row datasets via stratified sub-sampling, achieving a 10% reduction in RMSE.
- Tuned hyperparameters and selected regularization with cross-validation and grid-search, corrected prediction probability for imbalanced testing data, and estimated multiple performance metrics.

Alibaba Beijing, China

Data Scientist Intern, Algorithms *May 2022 - Sep 2022*

Alibaba is a top-30 global tech giant with 125,000+ employees.

- Developed and optimized a **graph-based routing framework** that explicitly modeled highway toll penalties and user cost sensitivity, integrating traffic signals and road network constraints, reducing toll-related customer complaints by 15%.
- Built ETA prediction models using **Gradient Boosted Trees** and **LSTM-based sequence models** to capture spatiotemporal traffic dynamics, integrating features from funnel analysis, improving ETA accuracy by ~12–15% in peak-hour scenarios.
- Designed and analyzed A/B experiments to evaluate an OCR-based vehicle identification feature, enabling riders to locate vehicles more quickly and reducing average pickup time by 20%, improving user experience significantly.

SELECTED PUBLICATIONS & CONFERENCE

[25' TGRS, IF: 8.6] H. Wang, F. Lei, A condition diffusion model for water retrieval incorporating terrain-score DEM

[25' RSE, IF: 11.4] H. Wang, F. Lei, E. N. Anagnostou, W. T. Crow, Advancing Root Zone Soil Moisture Time Series Prediction Using Deep Learning with Evapotranspiration and Microwave Soil Moisture

[24' JoH, IF: 6.3] H. Wang, F. Lei, Flood inundation mapping with CYGNSS: a Machine-learning-based framework