Marc Berghouse

650-576-4226 | marc.berghouse@gmail.com | Github | LinkedIn
Portfolio: Website | Google Scholar

Data Science | Hydrology | Machine Learning

Technical Skills

Programming Languages: Python, Matlab, R, SQL, Fortran, Bash, Stata
Machine/Deep Learning Frameworks: PyTorch, Tensorflow, Keras, Fastai, Jax, PyTorch Lightning, Torchvision, HuggingFace, TIMM, Sagemaker, Scikit-learn, Onnx, PySpark, XGBoost, LightGBM, Albumentations
Python Libraries: Pandas, Numpy, Matplotlib, Seaborn, Polars, Dask, Scipy, Requests, WandB, Json
Computer Vision: OpenCV, Classification, Object Detection, Object Tracking, Segmentation, Structure From Motion, Depth Estimation, Signal Processing, Biological/Medical Image Analysis, Pose Estimation
Machine Learning and Data Science: Time Series Analysis, Feature Engineering, Supervised Learning, Unsupervised Learning, Semi-Supervised Learning, Data Visualization, Scientific Data Analysis
Natural Language Processing: NLTK, spaCy, NetworkX, Sentiment Analysis, Recommendation Systems
Large Language Models (LLMs): RAG, Finetuning, Prompt Engineering, RLHF, QLoRA, LangChain
Software Development: Docker, Linux, Github, Bitbucket, Streamlit, REST APIs, AWS, Google Cloud, Django
General Computing: High Performance Computing, Blender, Kaggle, Simulated Data Generation, Multi-physics Simulations, CFD, Particle Tracking, Google Earth Engine, GIS, SLURM, Mathematical/Statistical Modeling
Lab Work: Biological Sample Prep, Microfluidic Device Prep, Microscopic Imaging of Bacteria, ICP-MS

Research Experience

University of Nevada, Reno, Graduate Program of Hydrologic Science   Reno, NV
Graduate Research Assistant, Prof. Rishi Parashar   *2020 – 2024*

- Developed DeepTrackStat – an accurate (less than 5% MAPE) PyTorch-based model that uses transfer learning with Volo-D3 to predict speed, turn angle, and dispersion statistics from a variety of image sequences of particles. (Code)
- Developed DeepRTUpscale – a PyTorch-based deep learning model with a novel architecture that can be used to upscale reactive transport simulations with 1% MAPE to achieve a 30x speedup. (Code) (Poster)
- Used Python, Matlab, Blender, Linux, Docker, Bash, Slurm, and Singularity to generate and analyze high-resolution CFD simulations of experimental microfluidic geometries on the UNR supercomputer. (Code)
- Recorded/Analyzed over 4 TB of high-resolution videos of bacteria in microfluidic devices using Python. (Code) (Video)
- Developed a novel background subtraction algorithm in Matlab to improve bacterial tracking capabilities. (Code)
- Developed novel statistical methods to investigate bacterial transport and particle tracking performance. (Code)
- Modified the PFLOTRAN source code (written in Fortran) to allow for velocity-based decay of biomass. (Code)
- Awarded $25,000 for the George Burke Maxey Fellowship in 2021 for my research proposal titled "*Integrating Microbial Motility Dynamics with Biogeochemical Reactive Transport Models: Applications to Bioremediation and Bioclogging*".
- Created and taught GEOL 701T, a graduate class at UNR on the applications of machine learning in hydrology. (Code)

University of Nevada, Reno, Department of Computer Science   Reno, NV
Graduate Research Assistant, Prof. George Bebis   *2020 – 2024*

- Developed the only publicly available implementation of the CvAM model for mammogram classification. (Code)
- Developed novel CNN architectures that are SOTA for whole-image mammogram classification. (Code) (Paper)
- Created a one-script framework for training and testing any TIMM model with a replaced classification head. (Code)
- Adapted TIMM source code for insertion of attention modules into CNN and transformer-based baseline architectures.
- Trained/tested more than 100 PyTorch vision model variations (2000 models in total) using WandB to track results.
- Developed a LayerCAM analysis script for the creation of > 100,000 activation heatmaps for over 2000 models. (Code)

SLAC National Accelerator Laboratory   Stanford, CA
Research Assistant, Profs. Zach Perzan, Kate Maher, and John Barger   *2018 – 2021*

- Designed a data processing pipeline to extract chemical concentrations and statistical trends from ICP-MS data.
- Combined data from multiple sources to curate a high-quality multi-site hydro-climatological time series. (Code)
- Used machine and deep learning models (including a novel physics-guided neural network) to predict soil moisture values with up to 95% accuracy and soil CO_2 values with up to 75% accuracy. (Code)

Columbia University, Department of Earth Science   **New York, NY**
Undergraduate Thesis, Prof. Blair Goodridge   *2014 – 2018*
- Used Python to extract water quality data from the USGS-NWIS REST API.
- Used Python and Stata to analyze statistical trends in water quality data via multiple linear regression. ([Thesis](#))

Side Projects

HydroML – A framework hosted on streamlit designed to be to an all-in-one application where hydrologists can develop machine learning and deep learning models | *PyTorch, Scikit-Learn, Streamlit* | [Code](#)
CodeGraph – A tool used to generate knowledge graphs of code repositories | *spaCy, NLTK, NetworkX* | [Code](#)
SkinScanner – A tool that combines standard image classification models with Mistral-7b to give users accurate information about skin, such as cancer presence, skin type, and skin health | *Django, LangChain, PyTorch* | [Code](#)
CodeBaseChatter – Tools for fine tuning open source LLMs on code scientific codebases | *Python, QLoRA* | [Code](#)

Publications

- Berghouse, M., Bebis, G., Tavakkoli, A. (2023). Investigating the Impact of Attention on Mammogram Classification. In: Bebis, G., *et al.* Advances in Visual Computing. ISVC 2023. Lecture Notes in Computer Science, vol 14361. Springer, Cham. ([Code](#))

Working Papers

- Revise & Resubmit at *Soft Matter* – Berghouse et al., "Advection-Dominated Transport Dynamics for Pili and Flagella-Mediated Motile Bacteria in Porous Media". ([Code](#)) ([Paper](#))
- Under Review at *Scientific Reports* – Berghouse et al., "Evaluation of Particle Tracking Codes for Dispersing Particles in Porous Media". ([Code](#)) ([Paper](#))
- Major Revision at *Image and Vision Computing* – Berghouse et al., "Exploring the Influence of Attention for Whole Image Mammogram Classification". ([Code](#)) ([Paper](#))
- In development – Berghouse et al., "DeepTrackStat - a deep learning model for the extraction of statistics from particle tracking videos". ([Code](#)) (Paper Draft Coming Soon)
- In development – Berghouse et al., "From Physical Insight to Predictive Modeling: Abiotic Feedbacks and Machine Learning for Biomass Growth in the Hyporheic Zone". ([Code](#)) ([Paper Draft](#))
- In development – Berghouse et al., "Impacts of Microbial Motility on Column-Scale Reactive Transport Simulations of U and V"

Conference Presentations

2021	Impacts of Flow on Transport of Motile Microbes in Synthetic Porous Media – American Geophysical Union, New Orleans
	Indirectly Measuring Soil Carbon Dioxide with Machine Learning – Knowledge Guided Machine Learning, Online
2022	Evaluation of Particle Tracking Codes for Dispersing Particles in Porous Media – American Geophysical Union, Chicago
	Modeling Hotspot and Preferential Flow Path Formation in the Hyporheic Zone – American Geophysical Union, Chicago
	Hydrodynamic Influences on the Transport of Motile Bacteria in Porous Media – Gordon Research Conference, Switzerland
2023	Investigating the Impact of Attention on Mammogram Classification – ISVC, Lake Tahoe
	Influence of abiotic factors on biomass growth in the hyporheic zone – American Geophysical Union, San Francisco
2024	Upscaling Reactive Transport Simulations With Deep Learning – UNR Data Science Initiative, Reno

Education

University of Nevada, Reno | PhD in Hydrology | 2020 – 2024 | 3.9/4.0
- Teaching Assistant – GE 614: Hydrologic Fluid Dynamics (2 semesters)
- Coursework – Computational Physics, Mammogram Detection with Deep Learning, Mathematical Modeling, Numerical Methods in Earth Science and Engineering, Fluid Dynamics, Environmental Chemical Kinetics

Columbia University | BA in Earth Science | 2014-2018
- Coursework – Big Data with Python, Calculus 3, Differential Equations, Linear Algebra, Statistics with Calculus, Econometrics, Physics with Calculus 1, 2, and 3, Chemistry 1 and 2, Atmospheric Chemistry, Physical Oceanography, Geochemistry