Alexandru Mihai

Research Data Scientist Seeking Industry Transition

+81 (70) 1641-4424 | contact@alexandru-mihai.com | alexandru-mihai.com | github.com/alexandru-mihai-94

Professional Summary

Research Data Scientist with a Ph.D. in Applied Mathematics and 8+ years of experience building computational solutions across diverse scientific domains at an international research institute in Japan. I serve as the sole data scientist for a marine climate change research unit, independently designing and delivering ML pipelines, satellite imagery analysis workflows, computer vision systems, NLP tools, and data infrastructure that support active research and publication. My work spans the full project lifecycle – from stakeholder consultation and problem definition through implementation, validation, documentation, and end-user training. I am seeking to transition to industry, where I can apply my breadth of technical skills, rapid domain adaptability, and experience delivering working solutions to product-oriented teams.

Technical Skills

Programming: Python (8+ years), R, SQL, C/C++
Machine Learning & Deep Learning: PyTorch, TensorFlow, CNNs (DenseNet, VGG, ResNet), transfer learning, UMAP, HDBSCAN, k-means, classification, clustering, model calibration, validation, uncertainty quantification
Computer Vision: 2D/3D image segmentation (Cellpose), object classification, image tiling, feature extraction, morphological analysis, automated image processing pipelines
NLP & LLMs: LLM-based text embeddings, sentiment analysis, TF-IDF, semantic network analysis, uncertainty estimation for LLMs (Mistral-7B, Llama, Phi-3 Mini)
Satellite & Geospatial: Google Earth Engine, Sentinel-2, LIDAR point cloud processing, drone imagery, change detection (MAD, PCA, band ratio, Otsu thresholding), QGIS, PostGIS, SpatiaLite, GDAL, Leaflet
Data Engineering: PostgreSQL, large-scale datasets (100TB+), ETL pipelines, data validation frameworks, SpatiaLite, data format conversion
Infrastructure: Docker, Linux (daily use), Git/GitHub, SLURM/HPC job scheduling, shell scripting, Jupyter, web development (JavaScript, HTML/CSS)
Other: 3D printing (design and fabrication), scientific writing, LaTeX, technical documentation

Professional Experience

Senior Research Data Scientist 2022 – Present
Okinawa Institute of Science and Technology (OIST) – Marine Climate Change Unit *Japan*

Sole data scientist for a research unit, responsible for all computational analysis, ML development, data infrastructure, and technical support across a portfolio of concurrent projects. Officially classified as a Research Unit Technician (bioinformatics), but operating scope extends across data science, software engineering, geospatial analysis, hardware fabrication, and fieldwork support.

- **Machine Learning & Computer Vision:** Designed and developed DEEPFIN, an unsupervised deep learning pipeline for fish image classification using CNN feature extraction, UMAP dimensionality reduction, and HDBSCAN clustering. Validated on clownfish developmental staging and the Fish4Knowledge benchmark (ARI: 0.873). Built and released as a pip-installable Python package with CLI interface. Manuscript in review at *Methods in Ecology and Evolution*.
- **Image Classification at Scale:** Developed LeafMosaic, a tile-based CNN pipeline classifying four aquatic macrophyte species (*Lemna minor*, *Spirodela polyrhiza*, *Azolla filiculoides*, *Salvinia natans*) using DenseNet-121, VGG-16, and a custom CNN with transfer learning. Achieved 91.8% mean validation accuracy across 19,754 tiles and 8 classification classes from a 69-unit mesocosm experiment. Software published on Zenodo; associated ecology manuscript in preparation.
- **Satellite Remote Sensing & Blue Carbon:** Lead satellite imagery analysis using Google Earth Engine for mangrove ecosystem monitoring in Okinawa. Integrating Sentinel-2 optical data, LIDAR point clouds, and drone imagery for change detection, biomass estimation, and carbon sequestration quantification. Separately built an end-to-end Sentinel-2 change detection pipeline implementing six methods (MAD, PCA, band ratio, Euclidean distance, image differencing, Green-Red Index) with seam correction, Otsu thresholding, polygon vectorization, and Leaflet visualization.
- **NLP & Biocultural Analysis:** Applied LLM-based text embeddings and sentiment analysis to anthropological interview data from Okinawan fishing communities (Ikei Jima, Itoman, Onna Village), extracting biocultural indicators related to marine ecosystem health. Built semantic network visualizations comparing community-level differences using TF-IDF and sentence embeddings.
- **LLM Uncertainty Research:** Developed training-free confidence scoring systems using hidden layer activations across multiple transformer model families (Mistral-7B, Llama variants, Phi-3 Mini) with evaluation across temperature settings.
- **Microscopy & Genomics:** Implemented 3D cell segmentation using Cellpose for z-stack microscopy images. Developed bioinformatics pipelines for fish genome methylation studies. Evaluated HPC versus workstation requirements for genomics workflows.
- **Coral Bleaching Analysis:** Built interactive Python and JavaScript visualizations for K-means cluster color data and CoralWatch bleaching scores across control/treatment mesocosm groups, including a custom bleaching index and multi-panel comparative figures.
- **Data Visualization & Communication:** Developed and maintained interactive web-based visualization platforms for eDNA and ecological data. Created data visualizations for stakeholder engagement and collaborative research interpretation.

- **Infrastructure & IT:** Manage on-site backup servers, cluster storage (100TB+ flash and bucket systems), and HPC infrastructure (SLURM). Provide ongoing technical support to lab members including software troubleshooting, data storage management, and system optimization. Maintain unit webpage.
- **3D Printing & Fabrication:** Design and produce custom research equipment including passive eDNA samplers for marine field deployment and reusable wet lab tools, reducing costs and enabling rapid prototyping.
- **Mentoring & Outreach:** Provide direct training and guidance to students and researchers on Python, deep learning frameworks, remote sensing data processing, and statistical methods. Conduct science outreach presentations to school children on nearby islands.
- **Fieldwork:** Participate in field operations including water collection for eDNA sampling and support for fish tracking experiments at the OIST marine station.

Research Scientist / Ph.D. Candidate 2017 – 2022
Okinawa Institute of Science and Technology (OIST) *Japan*

Investigated self-assembly systems within soap films, independently managing the full research pipeline from question discovery and experimental design through prototyping, physical data collection, imaging, image processing, quantitative analysis, and scientific writing.

- Developed mathematical and computational models for interfacial fluid dynamics and emergent geometric phenomena using optimization and statistical inference
- Designed and built physical experimental setups; wrote data acquisition software interfacing with cameras and sensors for automated high-throughput characterization
- Implemented custom image processing and analysis algorithms from scratch for experimental measurement extraction
- Applied rigorous statistical analysis and experimental design to large datasets; published and presented findings at international conferences

Teaching Assistant – Computational Mathematics 2020
Summer@ICERM, Brown University *Remote*

- Mentored student teams on computational mathematics topics including graph algorithms, spectral clustering, and iterative optimization methods
- Reviewed Python code architecture and guided reproducible research practices

Research Assistant – Medical Image Processing 2014 – 2015
Fraunhofer Institute for Digital Medicine MEVIS *Germany*

- Developed image processing and analysis algorithms for biomedical imaging datasets in a clinical research environment
- Completed bachelor's thesis on blood flow dynamics simulations in collaboration with MEVIS researchers

Publications & Open Source Software

Mihai, A., et al. "DEEPFIN: A Deep Learning Tool for Fish Image Classification from Unlabeled Data." *iScience Cell.* **In review**.

Mihai, A., et al. [Aquatic macrophyte mesocosm classification using LeafMosaic pipeline]. **In preparation**.

DEEPFIN – Open source Python package for unsupervised fish image classification. GitHub

LeafMosaic – Automated image classification pipeline for aquatic plant species. Zenodo DOI: 10.5281/zenodo.20421921

Education

Ph.D. Applied Mathematics & Computational Science 2017 – 2022
Okinawa Institute of Science and Technology (OIST) *Japan*
Thesis: Soap film mediated 3D Self-Assembly: Suspended and Displacement Driven Geometries using centimeter-scale tiles

B.S. Mathematics 2012 – 2015
Jacobs University Bremen *Germany*
Concentration: Statistics, Applied Mathematics, and Computer Science with biomedical applications

Language Proficiency

English: Fluent (native-level, professional working language) **Japanese**: Conversational (living in Japan since 2017)
Romanian: Native **French, German**: Conversational

Additional Information

Location: Based in Okinawa, Japan since 2017. Highly Skilled Professional visa holder.
Adaptability: Have contributed to projects spanning applied mathematics, fluid dynamics, marine biology, genomics, microscopy, remote sensing, NLP, and LLM research – each requiring rapid acquisition of new domain knowledge.
Community: Active in local science outreach and community engagement in Okinawa.