Muhammad Usman

Bruchwiesenanlage 4, 66125, Saarbruecken, Germany | +49 174 276 7793 |
usmanchaudhary.eng@gmail.com | LinkedIn | GitHub

Professional Summary

Bioinformatician and **DevOps-oriented Machine Learning Engineer** with expertise building scalable solutions in **genomics, transcriptomics, and proteomics**. In the final phase of a Master's degree exploring opportunities at the intersection of biological data analysis, machine learning, and infrastructure. Focused on building reproducible, scalable pipelines that translate complex biological data into insight.

Education

Bioinformatics, M.Sc. Oct 2022 – Present
Saarland University, Germany
- Thesis: *ML-Driven Exploration of NRPS Engineering in E. coli*
- Focus: Computational Biology, Genomics, NGS, Machine Learning

Biomedical Engineering, B.Sc. Oct 2018 – Aug 2022
University of Engineering and Technology (UET), Lahore, Pakistan
- Graduated with Honors (GPA: 3.72/4.00)
- Thesis: *ECG Analysis: Monitoring of Signs & Symptoms*

Work Experience

AI Safety Research Fellow Nov 2025 – Present
Saarland University, Germany
- Developing **multi-omics ML models** for personalized oncology using fusion and graph-based approaches.
- Assessing **uncertainty, causality, and robustness** under cohort and distribution shifts.
- Building reproducible **Snakemake + Docker** pipelines with AI-safety methods for reliable benchmarking.

Machine Learning & Bioinformatics Engineer Jun 2024 – Present
Myria Biosciences, Switzerland
- Processed and analyzed **10,000+ genomic and proteomic samples**, reducing pipeline runtime by **35%** using workflow managers and HPC clusters.
- Benchmarked variant calling pipelines, improving SNP detection sensitivity by **8%** against reference datasets.
- Designed and automated **ETL, NGS, and ML pipelines** (**Python, Pandas, SQL, Docker, CI/CD**), ingesting and validating **50 GB/week** of multi-omics data with **<2% error rate**.
- Built a **graph-based knowledge system (Neo4j)** integrating **10,000+ biological entities**, improving cross-domain query performance by **50%**.
- Delivered **30+ interactive dashboards and reproducible analytical reports**, version-controlled and containerized with **Git and Docker**, supporting decisions across **3 clinical programs**.

Research Assistant Oct 2024 – Jul 2025
Saarland University, Germany
- Applied **AI-driven CADD methodologies** for biosynthetic pathway optimization and molecule screening.
- Integrated synthetic biology with computational chemistry workflows for rapid in-silico experimentation.
- Automated ML pipelines using **Python, PyTorch, and Scikit-learn** on Linux, improving model testing and reproducibility.

Data Scientist in Drug Bioinformatics Jun 2024 – Dec 2024
Helmholtz Institute (HIPS), Germany
- Integrated NGS and **LC-MS/MS proteomics** data, enabling discovery of **15 novel BGCs**.
- Developed **machine learning models** improving expression-efficiency prediction accuracy by **22%**.
- Applied **graph-based methods** for pathway discovery, with **experimentally validated predictions**.
- Built version-controlled, reproducible **Snakemake pipelines** adopted by multiple research groups, saving **40+ hours/month**.

Research Assistant Dec 2022 – Sep 2024
Korean Institute of Science & Technology (KIST) Europe, Germany
- Supported sequencing-driven synthetic biology projects, validating **100+ cloning experiments**.
- Led student training and workflow standardization, reducing onboarding time by **30%**.
- Executed and analyzed **ELISA, cloning workflows, and phage–host interaction** studies to support bacteriophage engineering.

Volunteering

Research Mentor, Computational Biology Sep 2022 – Present
UET Lahore, Pakistan

- Mentored undergraduate research projects applying **machine learning and deep learning** to biological and biomedical data (EEG, genomic, physiological).
- Guided students in research design, data preprocessing, feature engineering, and model evaluation, strengthening scientific rigor and outcomes.
- Provided hands-on technical guidance in **Python, scikit-learn, PyTorch**, emphasizing reproducible and research-grade coding practices.

Projects & Certifications

- **GeneCoViz**, Co-expression network visualization (Plotly)
- **NeuroVision Explorer**, MRI analysis using CNNs/transformers
- **Heart Disease Prediction**, **92% accuracy** (RF/XGBoost)
- **Neo4j Knowledge Graph**, 10k+ biological entities
- **FastAPI CI/CD Reference**, Automated Docker deployment
- Neo4j Certified Professional; Graph Data Science Certified
- AWS Cloud Practitioner Essentials

Publications

Reprogramming Filamentous FD Viruses to Capture Copper Ions. *ChemBioChem.* DOI: 10.1002/cbic.202400237 2024

Bacteriophage Engineering for Improved Copper Ion Binding. *Macromolecular Bioscience.* DOI: 10.1002/mabi.202300354 2023

Electrocardiogram Data Visualization and Dimensionality Reduction. *ICBHS 2022* 2022

Brain Tumor Segmentation and Detection on MRI Images. *ICBHS 2022* 2022

Awards & Hackathons

Winner, Health Hack Saar Oct 2025
Saarbrücken, Germany
MEDPlex: Bridging the Gender Data Gap in Pediatrics

- Built a real-time **evidence-based pediatric medication recommendation system** addressing gender bias in dosing.
- Delivered a full working prototype within 48 hours with potential to improve medication safety and reduce dosing errors.

Finalist, ODDO BHF Equity Research Hackathon Nov 2025
Paris, France
Enhancing Equity Research Client Engagement

- Selected among top teams and invited to pitch at **ODDO BHF Headquarters** in Paris.
- Built an AI-augmented platform improving analyst–client engagement through availability signalling and automated follow-ups.

Participant, Defensive Acceleration Hackathon Nov 2025
Global
BioCast AI: Global Disease Outbreak Forecasting

- Developed a hybrid **LSTM + XGBoost** epidemiological forecasting pipeline with PCA/clustering analytics and outbreak risk prediction.

Skills

NGS & Bioinformatics: WGS, WES, RNA-seq (bulk & single-cell), methylome, variant calling, annotation, Bioconductor
ML & Data Science: Scikit-learn, PyTorch, PCA/UMAP/t-SNE, clustering, Random Forest, SVM, XGBoost, SHAP, GNNExplainer, graph ML
DevOps & Infrastructure: Docker, Git, Bash, Linux Server Management, CI/CD, Workflow Automation, Cloud Deployment
Pipelines & Tools: Snakemake, Nextflow, BioPython, Pandas, NumPy, Jupyter, Flask
Data Engineering & Databases: SQL, ETL/ELT, API pipelines, Neo4j knowledge graphs, data modeling
Visualization: Plotly, R Shiny, Jupyter Dashboards, RMarkdown, interactive reporting
Languages: English, German