

Amrithya Balaji

amrithya.github.io | amrithyarao1999@gmail.com | linkedin.com/in/amrithyab/ github.com/Amrithya

Summary

Graduate in MSc Machine Learning and Data Mining with experience across bioinformatics, NLP, and computer vision. Experienced in building interpretable deep learning models, with a background in data analysis, model optimization, and software development.

Education

MSc Machine Learning and Data Mining , University of Jean Monnet	Aug 2023 – Sep 2025
MSc Computer science ERASMUS+ semester , University of Freiburg	Oct 2024 – Mar 2025
B.E Computer Science and Engineering , SVCE, Chennai, India	Aug 2017 – May 2021

Experience

AI Researcher, Intern , IBDM, Marseille	April 2025 – Aug 2025
• Designing interpretable large language models (LLMs) to classify cellular states from bulk and single-cell RNA-seq data. Integrating explainability frameworks (e.g., SHAP, attention maps) to identify key genes influencing cell differentiation.	
• Outcome: improved classification accuracy and biological interpretability by >5% vs baseline models.	
Student Researcher, Part-Time , Bioinformatics Group, Albert-Ludwigs-University – Freiburg, Germany	Nov 2024 – March 2025
• Built a BERT and GRU based deep learning pipeline predicting neoantigen immunogenicity on experimentally validated datasets. Achieved 80% classification accuracy.	
Machine Learning Researcher, Intern , Laboratoire Hubert Curien – Saint Etienne	April 2024 – July 2024
• Enhanced graph learning models by incorporating screening-based sparsity methods, reducing runtime by 10–20%. Contributed to research on interpretable structured learning under varying optimization constraints.	
Programmer Analyst , Cognizant Technology Solutions – Chennai, India	July 2021 – Aug 2023
• Engineered and deployed full-stack web applications using Angular, JSP, and SQL Server for enterprise clients. Collaborated across 4 agile teams as DevOps Developer managing databases (Postgres, MongoDB) and build pipelines.	

Projects

Neoantigen Immunogenicity Prediction	– BERT & GRU-based classifier trained on 20k+ samples
Structured Generation from Call Transcripts using FastAPI and LLaMA	– Built an API to automatically generate structured summaries from customer call transcripts.
Transfer learning using ResNet	- Finetuned a pre-trained ResNet model on the MNIST dataset
Forest Cover Type Classification using SVM	- Implemented SVM achieving 85% accuracy
Performance Analysis of KNN on Waveform Dataset	- Benchmarked KNN under noisy conditions.

Skills

Programming Languages: C, C++, Java, Python, Angular
Data Science: Data Collection, Data Processing, Data Analysis, Data Visualization, Machine Learning, Artificial Intelligence, Natural language processing (NLP) – LLMs, Computer vision,
Deep Learning frameworks: PyTorch, Tensorflow, Keras
Frameworks & Libraries: OpenCV, Sci-kit learn, NumPy, Pandas, SciPy, Matplotlib
Tools: Git, MSOffice, Maven, Udeploy, Apache Subversion, Docker, Blender, HPC
Natural Languages: English: Fluent, French: Beginner, German: Beginner, Tamil: Native

References

M1 Internship supervisor: Benjamin Girault, Research Associate, INRIA E-mail: benjamin.girault@inria.fr
M2 Internship supervisor: Ronan Sicre, Université Paul Sabatier Toulouse III, Junior Professor Chair E-mail: ronan.sicre@lis-lab.fr
M2 Internship supervisor: Bianca Habermann, Group Leader Computational Biology, IBDM E-mail: bianca.HABERMANN@univ-amu.fr