Haricharan Bharathi

haricb03@gmail.com | github.com/Haricharan0311 | linkedin.com/in/haricharan-bharathi | Portfolio
+1 (984)255-5413 | 2512 Avent Ferry Road, Raleigh, North Carolina

Graduate student specializing in Machine Learning, with experience deploying LLMs and deep-learning pipelines at scale. Proven record of improving model accuracy and delivering measurable impact (e.g., $500K+ savings, top-10 global ranks).

EDUCATION

North Carolina State University, Raleigh, United States   Expected 2025
Candidate for Master of Science - Computer Science; CGPA: 4.66
Courses: Design and Analysis of Algorithms, Automated Learning and Data Analysis, Software Engineering, Computer Networks

Anna University, Chennai, India   Aug 2020- June 2024
Bachelor of Engineering - Computer Science and Engineering; CGPA: 9.675 , Silver Medalist
Courses: Data Structures, Operating Systems, Database Management Systems, GPU Computing, Compiler Design
Academic Achievements: Merit Scholarship awarded for highest CGPA - **Rank 2/140+** (Computer Science Department)

TECHNICAL FORTE

Languages: Python, C++, SQL, Bash, JavaScript
Tools/Frameworks: PyTorch, Tensorflow, Hugging Face, AWS, Docker, Git, BigTable, Tableau, React
Certifications: AWS - Certified Cloud Practitioner, Google Certified Data Analyst, NPTEL - Programming in Modern C++

WORK EXPERIENCE

Intern at Stanley Black & Decker, Towson MD - (Machine Learning Engineer)   June 2025 - Aug 2025
- Built a **zero-to-one drill malfunction prediction system** by processing raw IoT sensor data from field devices.
- Applied **feature engineering** and iterative tuning to boost prediction accuracy by **28%** over existing methods.
- Reduced **false warranty claims** by **$500K/year**, directly enhancing product reliability and operational efficiency.
- Deployed an **end-to-end data pipeline to AWS IoT**, enabling both local batch training and real-time failure predictions.

Intern at Indian Institute of Technology, Madras   Jan 2024 - June 2024
- Built a **deep-learning pipeline** for **Diabetic Retinopathy staging** using Group Equivariant Networks and ensembles.
- Achieved a **Kappa score of 0.95 (SOTA)**, with the system **deployed in 300+ hospitals across India**.
- **Benchmarked LLMs** (GPT-3.5, LLaMA-2, BioBERT) on medical corpora to evaluate performance in domain-specific tasks.
- Identified **tokenization strategies** that improved precision and guided future model selection for clinical NLP applications.

Intern at Tata Consultancy Services, Innovation Labs   Jul 2023 - Sept 2023
- Built an **LLM-based interface** for energy-forecast prediction, enabling domain experts to query and interpret forecasts.
- **Fine-tuned LLMs with PEFT and QLoRA** and integrated **tool calling to statistical methods**, creating a hybrid pipeline.
- Reduced **MAPE from 8.17 to 2.64**, significantly outperforming baseline statistical methods.
- Applied **prompt engineering**, crafting domain-specific prompts and integrating exogenous variables to boost accuracy.

PROJECTS AND PUBLICATIONS

Medical Caption Prediction (ImageCLEF) : Ranked 7th globally **-** Link to paper
- Tackled the challenge of **generating captions for clinical images** by detecting UMLS medical concepts as tags.
- Performed **concept detection using DenseNet121 and XGBoost**, extracting high-value clinical labels from imaging data.
- Utilized **LSTM networks with beam search** to generate context-aware captions from detected concepts and image features.
- Achieved **BERTScore 0.60 (benchmark 0.64)**, earning a **top-7 global rank** and **second-author publication-**ImageCLEF 2023.

Correlating Biomedical Image Fingerprints (ImageCLEF) : Ranked 9th globally - Link to paper
- Tackled **privacy and data-integrity risks** caused by synthetic GAN-generated images infiltrating biomedical datasets.
- Built a relational **ResNet-based few-shot network** with custom similarity metrics to detect such images.
- Validated with **clustering and SVM**s, earning a **first-author publication** at ImageCLEF 2023.

Advanced Hepatoma Staging (Machine Learning) :
- Adapted a private hospital dataset by **stain-color normalizing** against a Kaggle dataset to create a universal input standard.
- Trained a **U-Net** model to generate segmentation masks on the hospital data, enabling reliable region extraction.
- Used **CellProfiler (MIT open-source)** to extract key biomarkers from the segmented images for downstream staging.
- **Achieved** a **20% F1 score improvement** over baseline models, enhancing the accuracy of liver cancer staging.