SHASANKA SHEKHAR PADHI

shasankashekharpadhi@gmail.com | +917258949271 | Linkedin | Github | shasanksp.github.io | Bengaluru

SENIOR RESEARCH ASSOCIATE

SUMMARY

A Data Science/Machine Learning professional with around 3 years of experience in **AI/ML** model development, data pipelines, and knowledge graph applications. Skilled in **GenAI, NLP, LLM fine-tuning, and building knowledge graph solutions** to drive data-driven insights, enhance decision-making, and deliver scalable solutions through cross-functional collaboration across diverse industries.

Technical skills

Programming languages: Python | R | Shell Scripting | Cypher
AI/ML- Frameworks: PyTorch | Scikit-learn | TensorFlow | SciPy | NLP | LLM | OpenCV | GNN | GEN-AI | OLLama | LoRA | RAG
Deployment: Docker | GIT | Postman | AWS | MLflow | RESTAPI | LangChain
Databases: Neo4J | MongoDB | ChromaDB
Bioinformatics: NGS analysis [RNS-seq, scRNA-seq, WGS, WES] | GWAS | PK/PD Modeling

Work Experiences

Senior Research Associate — Jun '23 - Present

Syngene — Bengaluru

Received SPOT award for advancing bioinformatics workflow.

- **AI driven antibody sequence generation**: Developed an **AI** driven pipeline for de novo antibody sequence generation, integrated with sequence optimization and structural validation tools to design high affinity, **developable antibodies** with therapeutic potential.
- **Machine Learning models**: Built an auto ML pipeline and deployed in **AWS** server. This helped reducing dependency on data scientist for model building and increased capabilities across departments
- **Knowledge Graph**: Built a KG with open source biological/RWD data, harmonized with **controlled vocabularies** for each entity. Application included drug repurposing, target identification, safety assessment for toxicity and organ wise stratification, reducing months of work to weeks
- **Backend development and REST-API**: Designed **RESTful APIs** for an integrated drug discovery platform, providing data science features for the users, saving the data in **mongoDB**
- **Automated spectra processing tools**: Built tools for processing spectra data from screening experiments and assays to infer purity and yield for the compounds. This helped removing manual interventions accelerating drug discovery
- **Structure-based druggability**: Developed a pipeline to accelerate searches across a **vector database** containing binding pockets, enabling rapid identification of similar sites to assess target protein and possible adverse events
- ***In-silico* KO/perturbation**: Developed a high-throughput **Boolean model simulation** pipeline for *in-silico* gene knockout/perturbation experiments, supporting data-driven therapeutics and enhancing precision in target prioritization

Project Associate — Oct '22 - Jun '23

Centre for Brain Research, IISc — Bengaluru

- Standardizing pipeline for quality control of **GenomeIndia** GWAS WGS data by CBR IISc.
- Detection of positive selection in human populations through **cross-population studies**

Single Cell Curation Intern — Aug '22 - Oct '22

Elucidata — Remote

Curation and standardizing annotation for **scRNA seq** data.

Personal projects

- Built a **Deep learning** model (with **PyTorch**) using protein sequence embeddings for protein classification
- A **multi-output classification** model for breast cancer using iTRAQ proteome profiles of TCGA cancer samples and METABRIC mRNA levels
- Identified protein **biomarkers** that can discriminant between different experimental classes of mice with Down syndrome
- **TOX24 Challenge:** Predictive models for drug toxicity using data from TOX24 - **Graph Attention Network**
- **Image classification models** (**CNN**) using **MedMNIST** datasets for different modalities
- **Natural language processing**: Sentiment analysis for detecting mental health state
- **GenAI & RAG**: Developed a biomedical research assistant that streamlines literature exploration using **llama3** with a **chatbot** for natural language Q&A
- **CAMDA challenge:** Constructed a **Temporal Knowledge Graph** from diabetes patient records (**EHR**) using **Neo4j**, integrated with llama3 to setup a **RAG** workflow for various **AI driven medical applications**.

EDUCATION

Master of Technology in Bioinformatics

University of Hyderabad — Hyderabad

9.05 CGPA - GATE 2022 (AIR : 309)

Bachelor of Engineering in Biotechnology

Birla Institute of Technology — Ranchi

7.44 CGPA - GATE 2020 (AIR : 767)

Certifications

- Big, data, genes and medicine - The state university of New York (Coursera)
- Artificial Intelligence (Syngene)