SHASANKA SHEKHAR PADHI

✉ shasankashekharpadhi@gmail.com   📞 +917258949271   in [Linkedin](#)   ◯ [Github](#)   🌐 [shasanksp.github.io](#)   📍 Bengaluru

SENIOR RESEARCH ASSOCIATE

SUMMARY

A Bioinformatics/Machine Learning professional with 3 years of experience in AI/ML model development, biomarker identification, and knowledge graph applications. Skilled in **GenAI, LLM, building knowledge graph and bioinformatics solutions** to drive data driven insights, enhance decision making, and deliver scalable solutions through cross functional collaboration across diverse industries. Passionate about solving complex biomedical problems at the intersection of AI, omics, and healthcare.

Technical skills

Programming languages: Python | R | Cypher | Shell Scripting
AI/ML- Frameworks: PyTorch | Scikit-learn | TensorFlow | SciPy | LLM | GenAI | RAG
Deployment: Docker | GIT | RESTAPI | LangChain
Databases: Neo4J | MongoDB | ChromaDB
Bioinformatics: RNA seq, scRNA seq, WGS | GWAS

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 antibody sequence generation, integrated with sequence optimization and structural validation tools to design high affinity, **developable antibodies** with therapeutic potential.
- **Knowledge Graph**: Built a KG with open source biological/RWD/clinical 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.
- **Machine Learning models**: Developed an **automated ML** pipeline to build Quantitative Structure-Property Relationship (**QSPR**) model for drug property prediction, that helped reducing dependency on data scientist for model building and increased capabilities across departments.
- **GWAS - AllOfUs cohort**: Developed a Cromwell (WDL) pipeline to scale up association studies across 1000s of traits using **EHRs** in the cohort to identify **statistically significant** biomarkers.
- **Machine Learning driven biomarker identification:** Developed a computational workflow for biomarker identification, using **DEG analysis**, **WCGNA** and classical ML models with RNA seq data, to identify diagnostic and prognostic signatures, enabling patient stratification for precision medicine applications.
- *In-silico* **KO/perturbation**: Developed a high throughput **Boolean model simulation** pipeline for in silico gene knockout/perturbation experiments, using RNA seq data to initialize the system states supporting data driven therapeutics which enhances precision in target prioritization.
- **Structure-based druggability**: Developed a structure based druggability prediction pipeline leveraging parallel processing to accelerate searches across a database of known binding pockets, enabling rapid identification of similar sites to assess target protein druggability.

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.

Single Cell Curation Intern   Aug '22 - Oct '22
Elucidata   Remote

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

Personal projects

- **CAMDA challenge:** Constructed a **Temporal Knowledge Graph** from diabetes patient records (**EHR**) using **Neo4j**, integrated with llama3 to setup a **GraphRAG** workflow for various **AI driven medical applications**.
- **GenAI & RAG**: Developed a biomedical research assistant that streamlines literature exploration using **llama3** with a **chatbot** for natural language Q&A
- Built a **Deep learning** model (with **PyTorch**) using protein sequence embeddings for protein classification
- Built 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
- **QSPR Models - NeurIPS_Open Polymer Prediction Challenge:** Predictive models for polymer property prediction.

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)