Sravya Sri Mallampalli

Research Assistant

Sricity, Andhra Pradesh, India

📞 +91 6303457301 ✉ sravyasri.mallampalli@gmail.com 🔗linkedin.com/in/sravya-sri-mallampalli ⭘ github.com/Sravz2433

Education

Indian Institute of Information Technology, Sricity 2022 – Present
B.Tech in Computer Science and Engineering, CGPA: 6.91/10.0

Birla Public School, Doha, Qatar 2009 – 2022
Senior Secondary School Examination, CGPA: 8.6/10.0

Experience

Bioinformatics Research Intern April 2025 – Present
Shodhaka Life Sciences *Remote*

- Analyzing gene expression regulation by identifying hidden sequence features such as unknown motifs, CpG islands, and regulatory elements.
- Developing automation tools and scripts to streamline internal workflows for data curation and large-scale sequence processing.
- Contributing to the discovery and annotation of conserved non-coding genomic signals using custom bioinformatics pipelines.

Bioinformatics Research Intern 2024 – May 2025
IIIT Sri City, under the guidance of Dr. Chandramohan Dasari, Assistant Professor *Sricity, India*

- Built machine learning models (XGBoost, Random Forest) to classify *MECP2* gene variants associated with Rett Syndrome, achieving up to 97.5% accuracy.
- Processed and integrated genomic data from ClinVar, applying variant interpretation pipelines for pathogenicity assessment.
- Co-authored a workshop paper accepted at PM4B@PAKDD 2025.

Research Works

Thermodynamic Impact Prediction of pre-miRNA Mutations May 2025 – Present
⭘

- *Thermodynamic Impact Prediction of pre-miRNA Mutations.* Accepted for poster presentation at InCoB 2025.
- Developed an ANN-based model to predict thermodynamic changes ($\Delta\Delta G$) caused by point mutations in pre-miRNA hairpin structures.
- Automated mutant sequence generation and ΔG computation using ViennaRNA RNAfold, creating a curated dataset from miRBase.
- Tools and Libraries: TensorFlow, Keras, ViennaRNA, Biopython, Pandas, Matplotlib.

Pathogenicity Classification of MECP2 Mutations in Rett Syndrome Feb 2025 – September 2025
⭘ 🌐

- Built an ML/DL pipeline to classify *MECP2* variants as benign or pathogenic, supporting Rett Syndrome diagnostics.
- Engineered domain-specific features (e.g., trinucleotide context, mutation type); applied SMOTE and Optuna for data balancing and tuning.
- Benchmarked models including Logistic Regression, XGBoost, CatBoost, ANN, LSTM, and Autoencoder.
- Tools: Python, Scikit-learn, PyTorch, Optuna, Biopython, Pandas.

Genomic Variant Classification for MECP2 Mutation Pathogenicity 2025
⭘ ↗

- Sravya Sri Mallampalli, Chandra Mohan Dasari. *Genomic Variant Classification for MECP2 Mutation Pathogenicity.*
- Accepted at the PM4B 2025 workshop of the PAKDD 2025 conference.
- Classified genomic variants in the MECP2 gene associated with Rett syndrome using computational analysis.
- Tools and Software: XGBoost, Random Forest, One-Hot Encoding

Machine Learning & Deep Learning Projects

Protein Interface Prediction using Graph Neural Networks
Sept 2025 – Present

- Developing a GCN-based framework to predict protein–protein interaction interfaces by modeling residues as graph nodes and interactions as edges.
- Integrating ligand and receptor residue embeddings for pairwise classification, enabling scalable application to drug discovery and protein engineering.
- Tools and Software: PyTorch Geometric, Biopython, PDB, scikit-learn.

Depression Detection from Social Media Text
Jan 2025 - Present

- Designed a two-stage NLP pipeline using MentalBERT embeddings and ensemble models to detect early signs of depression from Twitter and Reddit posts.
- Achieved 82% accuracy, with improved classification of nuanced negative vs. mixed emotional expressions.
- Tools and Software: MentalBERT, scikit-learn, Transformers, Python.

FASTA to 3D Protein Structure Prediction
Jan 2025 - March 2025

- Developed a pipeline to predict 3D protein structures from FASTA sequences.
- Tools and Software: AlphaFold, MODELLER, PyMOL.

Computational Drug Discovery Regression Model with RF
Dec 2024

- Developed a Random Forest-based regression model to predict drug-target interactions for computational drug discovery. Trained on molecular descriptors and biological activity data to improve lead identification efficiency.
- Tools and Software: Scikit-learn, RDKit, Pandas, NumPy, Matplotlib.

Mouse Mammary RNA-Seq Analysis
April 2024

- Analyzed RNA-seq data from 12 mouse mammary tissue samples to identify differentially expressed genes and explore pathways relevant to breast cancer.
- Tools and Software: HISAT2, featureCounts, edgeR, DESeq2, ggplot2.

Software-Based Projects

Comprehensive Analysis of SARS-CoV-2
Oct 2024

- Conducted genome assembly, variant classification, and structural prediction of the SARS-CoV-2 spike protein.
- Tools and Software: SPAdes, Nextclade, PyMOL, AlphaFold.

DNA Sequence GC Content Calculator
2023

- Created a Python script to calculate GC content across large DNA sequences to study genome stability.
- Tools and Software: Biopython, Matplotlib, Pandas.

Certifications

Specialization in Bioinformatics(Coursera, UC SanDiego)

Genomic Data Science Specialization (Coursera, Johns Hopkins University)

Advance Bioinformatics Certification Course (Biogrademy)

Python for Data Science and AI (IBM)

Technical Skills

Programming Languages: Python, R, SQL, Bash, JavaScript, Java, MATLAB

Data Analysis and Machine Learning: Pandas, NumPy, SciPy, TensorFlow, Scikit-learn

Genomics Tools: BLAST, BWA, HISAT2, edgeR, MEGA

Visualization: ggplot2, Matplotlib, Cytoscape

Version Control and Cloud Platforms: Git, GitHub, AWS, Google Cloud