TARUN CHIKATIPALLI

Boston, MA | (857) 334-9665 | chikatipalli.t@northeastern.edu | linkedin.com/in/chikatipalli-tarun | https://github.com/TarunChikatipalli1

SUMMARY

Graduate student in Computational Biology and Bioinformatics with expertise in RNA-seq analysis, molecular docking, and machine learning-driven predictive modeling. Proficient in automating bioinformatics pipelines, integrating multi-omics data, and applying AI for drug discovery. Adept at transforming complex biological data into actionable insights to advance research. A collaborative researcher dedicated to impactful advancements in bioinformatics and biomedical data science.

EDUCATION

Northeastern University (Boston, MA, USA) **September 2023 – December 2025**

Master of Science in Bioinformatics.

Relevant coursework: Bioinformatics Programming, Introduction to Computational Methods in Bioinformatics, Statisticsfor Bioinformatics, Genomics in Bioinformatics, Transcriptomics in Bioinformatics, Collecting, Storing and Retrieve data.

Loyola Academy Degree and PG College, (Hyderabad, TS, India) **July 2020 - June 2023**

Bachelor of Science in Biotechnology

SKILLS

Programming Languages: Python, R and Bash

Bioinformatics Tools and Software: RNA-seq (HISAT2, Salmon, MultiQC, DESeq2, FastQC, featureCounts, Trimmomatic), scRNA-seq (Seurat, Scanpy, Monocle, CellRanger, UMAP, t-SNE), Variant Calling & Genomic Analysis (GATK, SAMtools, BCFtools, VCFtools, SnpEff, ANNOVAR, IGV), Pathway Analysis (GSEA, DAVID, ClusterProfiler, BioConductor), Drug Discovery (AutoDock Vina, iGEMDOCK, SWISS ADME, RDKit, Molecular Fingerprinting,), Structural Bioinformatics (PyMol, Chimera, Discovery Studio, Ramachandran Plot Analysis).

Workflow Automation & Cloud Computing: Nextflow, Bash Scripting, Docker, Slurm, Git, AWS, HPC, Conda.

Data Analysis and Machine Learning : RDKit, Scikit-learn, TensorFlow, PyTorch, QSAR Modeling, Clustering, PCA, ggplot2, Seaborn, Matplotlib, Plotly, corrplot, caret, Streamlit, Shiny (R).

WORK EXPERIENCE

Research Intern, Ciencia Labs | Hyderabad,India **January 2023– June 2023**

- Investigated the role of TMPRSS4, a cancer-causing gene, in various malignancies (e.g., colorectal, pancreatic, lung cancers) by identifying potential small molecule inhibitors to hinder its activity.,
- Performed molecular docking using iGEMDOCK and AutoDock Vina to evaluate binding affinities of 10,000 ligands sourced from the COCONUT database, filtered based on molecular weight (400–500 Da), solubility, and H-bonding properties.
- Validated ligand-receptor interactions by selecting top candidates with energy scores (-9 to -10 kcal/mol) and ensuring toxicity-free profiles through SWISS ADME.
- Constructed and analyzed TMPRSS4 structure using visualization tools (Discovery Studio) and evaluated structural quality via Ramachandran plots, identifying 10 high-affinity inhibitors with significant potential for cancer therapy.

PROJECTS

Interactive scRNA-seq Analysis Dashboard (Shiny, Seurat, Data Visualization, Bioinformatics) **November 2024 – January 2025**

- Developed a Shiny dashboard for single-cell RNA sequencing (scRNA-seq) data analysis using the Seurat package, enabling UMAP visualization, gene expression exploration, and metadata-based comparisons.
- Integrated features for uploading Seurat object files (.rds), generating UMAP plots by metadata columns, and creating feature plots for selected genes.
- Automated core functionalities such as file handling, metadata validation, and output generation, reducing manual steps by 50%, while enabling high-resolution plot downloads, reset options, and detailed metrics visualization for seamless data exploration.

Glioma Classification and Predictive Modeling (Data Analysis, Feature Engineering, Machine Learning) **August - October 2024**

- Conducted an in-depth analysis of a dataset containing 500+ glioma patients from the UCI Machine Learning Repository. The dataset included 24 features, comprising 4 clinical attributes and 20 mutation features to classify glioma grades (LGG and GBM) effectively.
- Enhanced data quality by addressing 10% missing values, removing 5% outliers (z-scores), normalizing features, and applying PCA to reduce dimensionality while retaining 95% variance. Designed and evaluated machine learning models (k-NN, Naïve Bayes, Decision Tree), achieving 97% accuracy with k-NN and validating performance using precision, recall, F1-score, and ROC-AUC.
- Visualized data patterns and model insights using tools like ggplot2, corrplot, and caret, ensuring interpretability and effective feature selection for improved model reliability.

Automated RNA-Seq Analysis Pipeline (Nextflow, Differential Gene Expression) **May 2024 – July 2024**

- Developed a Nextflow-based RNA-seq pipeline for analyzing differential gene expression across 10+ samples, cutting analysis time by 40%.
- Configured the computational environment with Conda and Docker, ensuring reproducibility and seamless dependency management. Installed essential bioinformatics tools (FastQC, HISAT2, SortMeRNA, featureCounts, MultiQC), managed dependencies to ensure reproducibility.
- Conducted data preprocessing steps: quality control (20-30% of low-quality reads removed), trimming, rRNA removal (achieving >95% filtering), sequence alignment (85-95% alignment rates), and gene count summarization.
- Automated the workflow using Nextflow, reducing manual intervention by 70%, and generated MultiQC reports across 5+ analysis stages, enabling efficient visualization. Performed DESeq2 analysis, identifying 1,000+ significantly expressed genes (FDR < 0.05) and 20+ enriched pathways for functional insights.

Bioactivity Prediction Web App (Machine Learning, Drug Discovery) **March 2024 – April 2024**

- Designed and Developed a Streamlit web app for real-time bioactivity prediction, improving accuracy by 30% through the integration of advanced machine learning algorithms.
- Managed a conda environment with over 20 libraries, including Pandas, NumPy, and TensorFlow, ensuring seamless integration and compatibility.
- Implemented a machine learning model, processing test data for 1,000+ compounds and generating a highly accurate model file.
- Launched an interactive app for real-time bioactivity prediction of 200+ compounds, enhancing data input, processing, and visualization while reducing analysis time by 50% and improving prediction accuracy by 30%.