Celestial BioML Model Hub

PEGASUS

Project Idea

Problem Statement:

– Space biology research lacks large datasets, making ML training difficult.

– Transfer learning can improve model accuracy and reduce training costs.

– We propose a Space Biology Model Zoo for pre-trained ML models tailored for space biology.

Our Solution

- Curate and organize biomedical datasets for pre-training.
- Identify space biology datasets from sources like NASA OSDR.
- Preprocess datasets and fine-tune ML models using transfer learning.
- Create a structured Model Hub for researchers to easily access these models.

Tech Stack

Programming & ML

- Python (Data Processing, Model Training)
- TensorFlow/PyTorch (Transfer Learning, Neural Networks)
- Pandas & NumPy (Data Handling & Processing)
- Matplotlib & Seaborn (Data Visualization)

Implementation Plan

Phase 1: Dataset Curation

- Collect publicly available biomedical datasets for pre-training models.
- Identify relevant space biology datasets from NASA OSDR.
- Structure data into an accessible format (CSV/Database).

Implementation Plan

Phase 2: Preprocessing & Model Training

- Normalize RNA sequencing & imaging data for ML models.
- Use pre-trained models and fine-tune on space biology datasets.
- Train a simple ML model and analyze performance.

THANK YOU