Technical & Teaching Portfolio

Beyza Kaya

beyzoskaya.github.io

This portfolio demonstrates my experience in computational biology, graph neural networks, machine learning for bioinformatics, and teaching assistantships.

Research Projects & Repositories

PGM-TL — Multi-task Learning with Protein Language Models

Repository: https://github.com/beyzoskaya/PGM-TL
  Contribution: Developing a multi-task learning framework built on protein language models to jointly optimize multiple downstream biological prediction tasks.

- Investigates **shared encoder representations** for protein sequences to improve performance and data efficiency across heterogeneous downstream objectives.

- Studies **task trade-offs, transfer effects, and representation reuse** under limited labeled data regimes.

- Explores principled task balancing strategies to avoid negative transfer while enhancing biologically meaningful representations.

- Designed as a flexible experimental framework to evaluate how different supervision signals shape learned protein embeddings.

- Ongoing research project with active development and experimentation.

DeepAllergen — Deep Learning for Non-homologous Allergen Identification

Repository: https://github.com/seferlab/deepallergen
  Role: Project Lead and Research Supervisor (Senior Design Project) **Contribution:** Supervised and guided a senior design project conducted by two Computer Science undergraduate students, defining the research direction and methodological framework for deep learning–based allergen identification.

- Formulated the core research problem, addressing limitations of homology-based allergen prediction methods in detecting **non-homologous allergens**.

- Guided architectural choices, experimental design, and evaluation strategy, including benchmarking against existing methods and conducting ablation studies.

- Mentored students throughout the full research cycle, including literature review, model development, result interpretation, and scientific reporting.

GraphVelo — Hybrid Geometric Deep Learning for RNA Velocity

Repository: https://github.com/beyzoskaya/graphvelo
  Contribution: Developed a Python library that integrates Graph Attention Networks (GATs)

with scVelo to denoise high-dimensional differentiation dynamics.

- Designed a **Hybrid Architecture** that fuses transcriptomic embeddings with biological priors (cell cycle scores, differentiation entropy) to resolve complex branching topologies.

- Achieved **90% cosine similarity** to ground truth dynamics in the Dentate Gyrus, significantly outperforming standard GCN baselines (87%).

- Built a custom ODE solver on the learned latent manifold to perform **computational lineage simulation** and a generative autoencoder for full gene expression reconstruction.

- **Software Engineering:** Structured as a modular Python library with `setup.py` installation, unit testing via `pytest`, and automated benchmarking scripts for reproducibility across datasets.

GAT-HiC — Graph Attention Model for Chromatin Structure Prediction

Repository: https://github.com/beyzoskaya/GAT-HiC
  Contribution: Developed a graph-attention and residual network-based model to predict three-dimensional chromosome structures from Hi-C interaction matrices.

- Combines unsupervised vertex embeddings (Node2vec) with attention-based graph neural networks to predict each genomic loci's 3D coordinates.

- Generalizes to dissimilar Hi-C datasets: different cell populations, restriction enzymes, and resolutions across human and mouse.

- Demonstrates improved accuracy of 3D chromosome shape reconstruction compared to existing approaches.

- Provides a robust framework for inferring genome-wide topological dynamics from interaction data.

STEP — Spatio-Temporal Expression Prediction

Repository: https://github.com/beyzoskaya/STEP
  Contribution: Developed neural network-based spatio-temporal graph models (STEPmr for mRNA, STEPmi for miRNA) to predict temporal gene expression dynamics.

- Integrates diverse Hi-C datasets and Hi-C–derived features to capture genome-wide interacting dynamics of genes.

- Predicts mRNA and miRNA expression with high accuracy: correlation of 77% and 93%, mean squared errors 0.21 and 0.01, explaining 59.1% and 88% of variance, respectively.

- Identifies genes with consistent high performance as structural signaling genes or transcriptional regulators with essential biological roles across tissues.

- Models dynamic regulatory interactions over multiple time points, linking chromatin architecture to temporal gene expression patterns.

BioinformaticML — Machine Learning Implementations for Bioinformatics

Repository: https://github.com/beyzoskaya/BioinformaticML
  Contribution: Implemented algorithms from the book *"Machine Learning for Bioinformatics, 2nd Edition"* by Pierre Baldi.

- Translated theoretical concepts from the book into hands-on code for deeper understanding.

- Covered supervised and unsupervised learning, feature selection, and classification algorithms.

- Provided a practical framework to explore bioinformatics problems and strengthen conceptual foundations.

DiffNFTGen — Generative Model for NFTs (Senior Design Project)

Repository: https://github.com/seferlab/diffnftgen
 Contribution: Developed a generative model for high-quality NFT images using Stable Diffusion and reinforcement learning-based rarity metrics.

- Collected specific NFT datasets from OpenSea to train the model on real-world distributions.

- Integrated a custom **RarityReward** function to quantitatively define and prioritize rare NFTs.

- Employed Proximal Policy Optimization (PPO) with Kullback-Leibler (KL) divergence regularization to balance rarity promotion with style fidelity.

- Ensured generated NFTs achieve uniqueness and high visual quality.

- Implemented as a Senior Design Project under Sefer Lab.

Teaching Assistantships

CS104 — Introduction to Programming (Python)

Repository: https://github.com/beyzoskaya/python-labs
 Contribution: Led lab sessions where students attempt coding exercises during the first hour and solutions are demonstrated live.

- Designed problem-solving exercises covering Python basics, loops, conditionals, and functions.

- Provides step-by-step demonstrations of solutions to ensure comprehension.

CS201 — Data Structures and Algorithms (Java & C++)

Repository: https://github.com/beyzoskaya/DataStructures
 Contribution: Assisted lab sessions on implementation and testing of data structures.

- Students practice linked lists, trees, graphs, and hash maps in Java and C++.

- Prepared test cases and debugging demonstrations to solidify understanding of algorithmic behavior.