Development of a Deep Learning-Based Drug-Target Interaction (DTI) Prediction Model for Precision Oncology

Project Description

Drug-Target Interaction (DTI) prediction plays a critical role in drug discovery, drug repurposing, and precision medicine. Identifying whether a drug binds to or modulates a specific protein target is essential for understanding therapeutic efficacy and potential side effects. Traditional experimental screening approaches are costly, time-consuming, and not scalable to the rapidly growing number of drug-like molecules and protein targets.

Recent advances in artificial intelligence (AI) and deep learning (DL) have enabled computational models to predict drug–target interactions using molecular structures and protein representations. However, building accurate and biologically meaningful DTI models requires careful data integration, representation learning, and evaluation strategies.

This project aims to develop a deep learning-based DTI prediction framework that integrates molecular graph representations of drugs with sequence- and structure-based protein embeddings. The model will be trained on curated public databases such as ChEMBL and BindingDB and will focus on human protein targets, with potential applications in cancer-related therapeutic discovery.

Students will contribute to dataset preparation, model architecture design, representation learning, training on GPU infrastructure, and performance evaluation. The project may later be extended toward cancer-specific drug repurposing analysis.

Duties / Activities
Students will:
- Work with large-scale molecular and protein datasets
- Perform data preprocessing and normalization
- Implement deep learning architectures in PyTorch
- Train and debug GPU-based models
- Evaluate model performance and interpret results
- Conduct ablation studies and hyperparameter tuning
- Document findings and prepare reports

Required Skills
- Strong programming skills in Python
- Understanding of machine learning fundamentals
- Familiarity with PyTorch or other deep learning frameworks
- Basic knowledge of data preprocessing and evaluation metrics
- Experience with Git and version control
- Ability to debug and work in GPU environments

<u>Nice to Have:</u>

- Experience with Graph Neural Networks
- Familiarity with PyTorch Geometric
- Understanding of molecular representations (SMILES, graphs)
- Basic knowledge of protein biology
- Experience working on HPC clusters

Learning Opportunities

Students will gain experience in:

- AI-driven drug discovery
- Graph Neural Networks
- Protein representation learning
- GPU model training
- Large-scale biomedical data engineering
- Translational AI applications in oncology

Potential outcomes include:

- Research publication contributions
- Extension into cancer-specific DTI modeling
- Development of a reusable DTI framework

Expected Team Size: *2-4*

Mentor

Name: Dr. Mohamed Elshrif   email: melshrif@hbku.edu.qa