Comprehensive Datasets for Graph Neural Networks in Drug Discovery

Abstract

We present a comprehensive collection of datasets specifically designed for training and evaluating Graph Neural Networks in drug discovery applications. These datasets cover various aspects of molecular property prediction and drug-target interaction.

1. Dataset Overview

Our dataset collection includes molecular structures, biological activities, and physicochemical properties. Each dataset is carefully curated and annotated for machine learning applications.

2. Molecular Property Datasets

We provide datasets for predicting molecular properties such as solubility, toxicity, and drug-likeness. These properties are crucial for drug development and safety assessment.

3. Drug-Target Interaction Datasets

Our interaction datasets contain experimentally validated drug-target pairs with associated binding affinities and inhibition constants. These datasets enable training of models for predicting novel drug-target interactions.