



Contents lists available at ScienceDirect

Life Sciences

journal homepage: www.elsevier.com/locate/lifescie

MPNN-CWExplainer: An enhanced deep learning framework for HIV drug bioactivity prediction with class-weighted loss and explainability[☆]

Aga Basit Iqbal^a, Assif Assad^a, Basharat Bhat^b, Muzafar A. Macha^c, Syed Zubair Ahmad Shah^{a,*}

^a Department of Computer Science and Engineering, Islamic University of Science and Technology, Awantipora, Jammu & Kashmir, India

^b Centre for Artificial Intelligence and Machine Learning, Sher-e-Kashmir University of Agricultural Sciences and Technology, Shalimar, Srinagar, Jammu & Kashmir, India

^c Watson-Crick Centre for Molecular Medicine, Islamic University of Science and Technology, Awantipora, Jammu & Kashmir, India

ARTICLE INFO

Keywords:

MPNN
GNNExplainer
Class-weighted loss
Molecular property prediction
Drug discovery
Explainable AI

ABSTRACT

Aims: Human Immunodeficiency Virus (HIV) remains a critical global health concern due to its impact on the immune system and its progression to Acquired Immunodeficiency Syndrome (AIDS) if untreated. While anti-retroviral therapy has advanced significantly, challenges such as drug resistance, adverse effects, and viral mutation necessitate the development of novel therapeutic strategies. This study aims to improve HIV bioactivity prediction and provide interpretable insights into molecular determinants influencing bioactivity.

Materials and methods: We propose MPNN-CWExplainer, a novel graph-based deep learning framework for molecular property prediction. The model integrates a Message Passing Neural Network (MPNN) with a class-weighted loss function to effectively address class imbalance in HIV datasets. Furthermore, GNNExplainer is incorporated to provide post-hoc interpretability by identifying key atom- and bond-level substructures contributing to model predictions. Model robustness is ensured through Bayesian hyperparameter optimization and multiple independent runs.

Key findings: MPNN-CWExplainer achieved state-of-the-art predictive performance on the HIV dataset, with an AUC-ROC of 87.631 % and AUC-PRC of 86.02 %, surpassing existing baseline models. The class-weighted approach enhanced minority class representation, and GNNExplainer successfully highlighted chemically meaningful substructures correlating with bioactivity.

Significance: The proposed framework not only improves prediction accuracy for HIV bioactivity but also enhances transparency and interpretability, crucial for medicinal chemists in understanding model behaviour. MPNN-CWExplainer serves as a robust and interpretable tool for computational drug discovery, supporting informed decision-making in lead optimization and molecular design.