

Encadrant : Warith EDDINE DJEDDI

Titre :

"Prédiction de l'Affinité de Liaison Protéine-Ligand à l'Aide des Modèles Transformers Préentraînés"

Description :

La prédiction de l'affinité de liaison entre une protéine et un ligand est une étape clé dans le processus de découverte de médicaments, permettant de réduire les coûts et le temps de développement des traitements. Bien que de nombreuses méthodes basées sur l'apprentissage profond aient été développées pour prédire cette affinité en exploitant les informations structurales ou séquentielles des médicaments et de leurs cibles, elles présentent plusieurs limitations.

Les approches basées uniquement sur les séquences peuvent ignorer des données essentielles, telles que les liaisons hydrogène, ce qui entraîne une perte d'informations importantes. À l'inverse, les méthodes basées sur les graphes risquent d'inclure des informations non pertinentes qui n'ont pas d'impact direct sur la prédiction. De plus, l'absence de partitionnement structuré peut limiter la capacité à représenter les caractéristiques clés des interactions.

Pour surmonter ces défis, l'étudiant doit proposer un modèle multimodal de prédiction de l'affinité de liaison protéine-ligand. Ce modèle intègre des Graph Neural Networks (GNN) ainsi que diverses représentations issues de modèles préentraînés pour les protéines et les ligands. En combinant des embeddings générés à partir de modèles Transformers avec des données structurées sous forme de graphes, notre approche capture les caractéristiques essentielles des interactions tout en évitant les informations superflues.

Les expérimentations menées sur deux jeux de données de référence, **KIBA** et **DAVIS**, montrent que le modèle proposé dépasse les performances des méthodes existantes. Cette recherche contribue à l'amélioration des outils de prédiction des interactions médicamenteuses, avec un fort potentiel pour accélérer le processus de découverte de nouveaux médicaments.

Objectif de la Recherche :

Développer un modèle prédictif avancé, basé sur des modèles Transformers préentraînés, pour améliorer la précision de la prédiction de l'affinité de liaison entre protéines et ligands en utilisant une approche multimodale qui combine les GNNs avec des techniques d'embedding modernes.

Mots-clés : Affinité Protéine-Ligand, Modèles Transformers, Graph Neural Networks, Prédiction Multimodale, KIBA, DAVIS, Deep Learning.

References :

[1] M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber, and P. P. Zarrinkar, "Comprehensive analysis of kinase inhibitor selectivity," *Nature biotechnology*, vol. 29, no. 11, pp. 1046–1051, 2011.

[2] J. Tang, A. Szwarzajda, S. Shakyawar, T. Xu, P. Hintsanen, K. Wennerberg, and T. Aittokallio, "Making sense of large-scale kinase inhibitor bioactivity data sets: a comparative and integrative analysis," *Journal of Chemical Information and Modeling*, vol. 54, no. 3, pp. 735–743, 2014

[3] <https://huggingface.co/microsoft/BiomedNLP-BiomedBERT-base-uncased-abstract-fulltext>

[4] <https://huggingface.co/jglaser/protein-ligand-mlp-3>

[5] WU, Hongjie, LIU, Junkai, JIANG, Tengsheng, et al. AttentionMGT-DTA: A multi-modal drug-target affinity prediction using graph transformer and attention mechanism. *Neural Networks*, 2024, vol. 169, p. 623-636.