

Autor

Kacper Stankiewicz

Promotor

dr. inż. Wojciech Czech

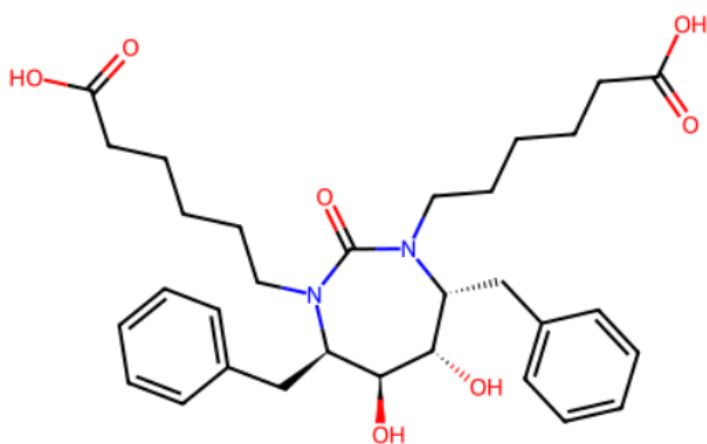
Temat pracy

Multimodal models for predicting drug-target interactions: integration of domain knowledge and deep learning

Opis

Celem pracy jest zaprojektowanie i implementacja multimodalnych modeli głębokiego uczenia maszynowego do przewidywania interakcji leków z celami molekularnymi (Drug-Target Interaction, DTI). Praca skupi się na integracji różnych źródeł danych, takich jak struktura chemiczna związków, dane sekwencyjne i strukturalne białek oraz profile ekspresji genów, w celu poprawy skuteczności predykcji DTI.

O[C@@H]1C@@HC@@HN(CCCCCC(=O)=O)C(=O)N(CCCCCC(=O)=O)[C@@H]1Cc1ccccc1



Sekwencja białek:

PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILIEICGHKAIGTVLVGPTP
VNIIGRNLLTQIGCTLNFPQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYDQILI
EICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF

Ki: 0.24

Źródła informacji:

- Struktura chemiczna związków (SMILES, grafy molekularne, embeddingi takie jak TopologicalTorsionFingerPrint, ECFPFingerprint)
- Dane sekwencyjne i strukturalne białek (sekwencje aminokwasowe, struktury 3D, embeddingi białkowe takie jak ProtBERT, ESM)
- Profile ekspresji genów i sygnatury transkryptomiczne (dane z inicjatyw takich jak LINCS)

Harmonogram (szacowany)

Etap	Opis	Termin rozpoczęcia	Termin zakończenia
Zebranie narzędzi	Zebranie i konfiguracja narzędzi (dostęp do zbioru danych, dostęp do superkomputera AGH)	01-11-2025	15-11-2025
Przegląd literatury i implementacja podstawowych modeli	Przegląd literatury dotyczącej multimodalnych modeli DTI oraz implementacja podstawowych architektur (GNN, CNN)	16-11-2025	31-01-2026
Implementacja zaawansowanych modeli	Implementacja multimodalnych modeli DTI z różnymi strategiami fuzji danych opartych na transformerach	01-02-2026	30-04-2026
Własne eksperymenty i rozwiązania	Przeprowadzenie eksperymentów porównawczych, analiza wyników oraz opracowanie własnych rozwiązań i ulepszeń	01-05-2026	31-05-2026
Ostatnie poprawki i przygotowanie pracy	Finalizacja pracy magisterskiej, ostatnie poprawki, podsumowanie, przygotowanie do obrony	01-06-2026	24-06-2026

Ocena ryzyka pracy

- Trudności z dostępem do zbiorów danych lub superkomputera AGH mogą opóźnić rozpoczęcie pracy nad modelem (Ryzyko: nisko-średnie, wpływ: ogromny)
- Trudność w implementacji zaawansowanych architektur multimodalnych (Ryzyko: małe, wpływ: duży)
- Niespojność danych bądź słabo zbalansowane zbiory danych mogą wpłynąć na jakość wyników eksperymentów (Ryzyko: średnie, wpływ: średni).

Ogólna struktura pracy

1. Wprowadzenie
2. Cel i zakres pracy
3. Przegląd literatury
4. Modele:
 - CNN, GNN
 - Transformer

- Multimodalne

5. Porównanie modeli i źródeł informacji
6. Eksperymenty własne
7. Wnioski i przyszłe kierunki badań
8. Bibliografia