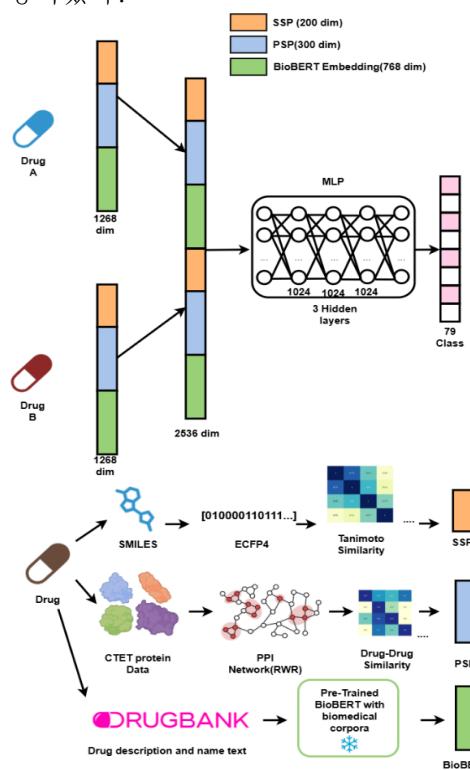




약물의 의미적 특성을 반영하기 위해, DrugBank의 약물 설명에 약물명을 결합한 텍스트 총 9,377개 약물에 대해 텍스트를 수집한 뒤, 특수문자 제거 등의 전처리를 수행했다. 이를 BioBERT 모델에 입력해 약물 단위 임베딩 벡터를 생성했고, 입력 크기가 512 토큰을 초과하는 경우는 제외했다. 최종적으로 9,294개의 약물에 대한 (1, 768)차원의 임베딩 벡터를 확보했다.

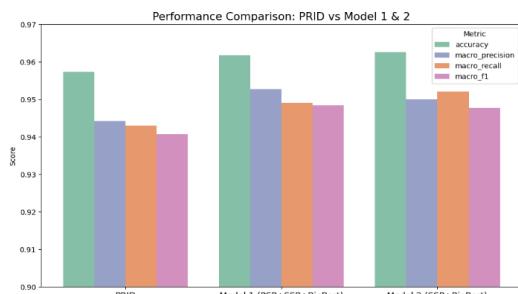
약물 간 화학 구조 유사도를 계산하기 위해, 약물별 SMILES 문자열을 기반으로 RDKit을 사용해 ECFP4를 생성했다. SMILES 정보는 주로 DrugBank에서 수집했고, 누락된 약물은 PubChem을 통해 보완했다. 총 13,082개 약물을 이용해 모든 약물 쌍에 대해 tanimoto 유사도를 계산했다. 본 연구에서는 이를 PCA를 통해 200차원으로 축소하여 모델 입력으로 활용하였다.



(그림 1) 전체적 모델 구성도. PSP는 PRID 연구에서 제안한 방식과 동일하게 구성했다. 약물과 연관된 CTET 단백질(carriers, transporters, enzymes, targets)을 STRING Database 기반 PPI 네트워크 상에 매핑하고, 이를 시드 노드로 설정하여 Random Walk with Restart(RWR) 알고리즘을 [3] 수행했다. 각 약물에 대해 계산된 RWR 전파 결과를 바탕으로 PSP 벡터를 생성해, 간접적인 생물학적 연관성을 반영한 표현으로 활용했다. 이를 PCA를 통해 300차원으로 축소하여 입력으로 사용했다. 세 가지 modality로부터 얻어진 벡터는 약물 단위로 결합되며, DDI 예측을 위해 3 개의 은닉층과 ReLU, Dropout 을 포함한 MLP 기반 다중 분류 모델이 최종

적으로 79 개의 DDI class에 대해 softmax 출력 값을 생성한다. 여러 modality의 조합으로 학습한 모델의 성능을 비교했다.

### 3. 결과



(그림 2) PRID 모델과 2 가지 조합의 모델들에 대한 지표 비교 그래프

두 가지 조합의 모델들은 PRID 모델 대비 정확도를 포함한 지표들에서 좋은 결과를 보였다. 이는 약물 간 의미적 정보를 반영한 표현이 모델의 예측 성능을 높이는 데 효과적임을 시사한다.

### 4. 결론 및 향후 연구 방향

본 연구에서는 DDI 예측을 위해 사전학습된 모델 BioBERT로 약물들의 관련 논문 기반 임베딩 정보를 얻었고, 이를 약물의 화학적 구조 유사도와 약물의 CTET 단백질간의 간접적 관계를 PPI 모델에 기반한 유사도를 함께 사용했다. 사전학습된 모델을 통한 약물의 특성을 반영함으로써 실제 DDI 예측 능력이 향상됨을 확인했다. 추후 DDI의 형태가 기술된 문장을 순수 PubMed 문헌만으로 사전학습된 언어모델인 PubMedBERT를 추가 학습(fine-tuning)하는 방안을 고려하고 있다. 이를 통해 도메인 특화 임베딩 표현을 확보하고, 보다 정밀한 약물 간 의미 정보 추출이 가능할 것으로 기대된다.

### 5. 사사

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2023R1A2C1007756).

### 참고문헌

- [1] Seo, J., Jung, H., Ko, Y. PRID: Prediction Model Using RWR for Interactions between Drugs. *Pharmaceutics*, 15(10), 2469, 2023.
- [2] Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res* 2018, 46, D1074–D1082.
- [3] Valdeolivas, A.; Tichit, L.; Navarro, C.; Perrin, S.; Odelin, G.; Levy, N.; Cau, P.; Remy, E.; Baudot, A. Random walk with restart on multiplex and heterogeneous biological networks. *Bioinformatics* 2019, 35, 497–505.