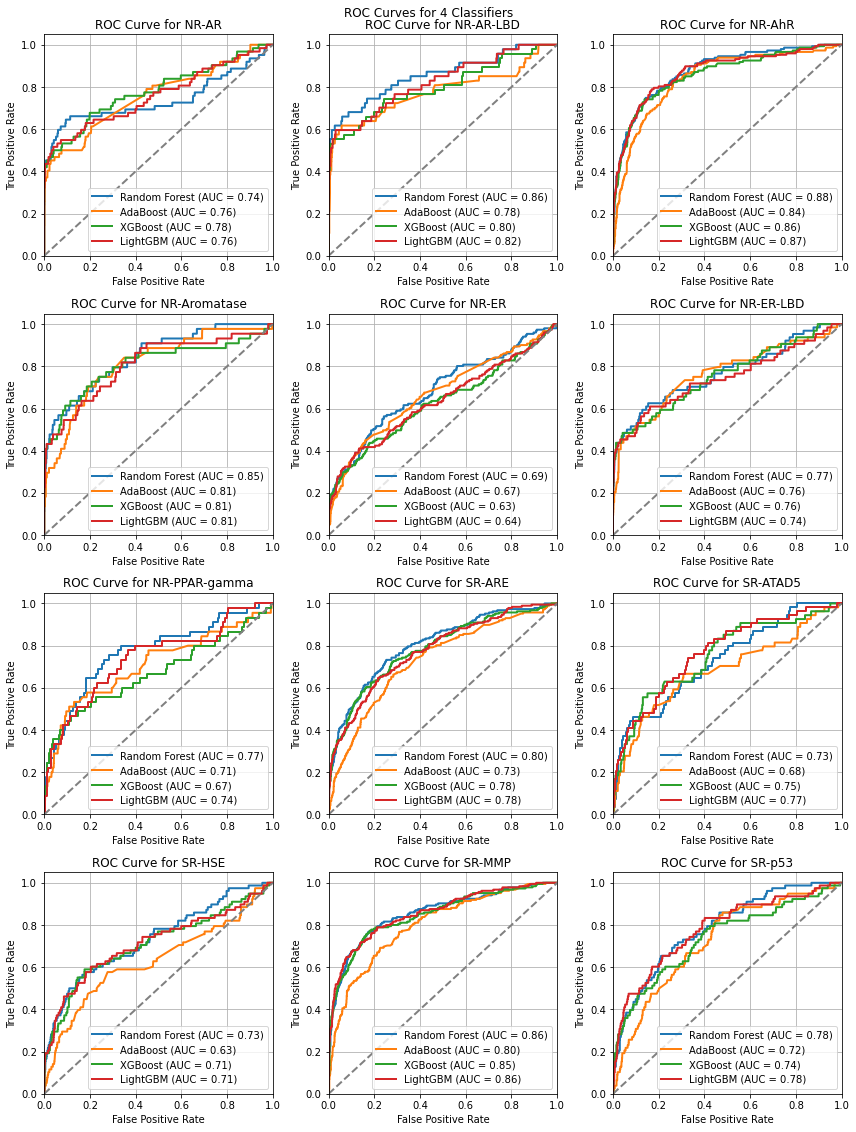
**Tox21 dataset 활용 독성예측모델**



Tox21 Dataset([tripod.nih.gov/tox21/challenge/data.jsp](https://tripod.nih.gov/tox21/challenge/data.jsp))을 이용, SMILES 데이터를 MolFinder 라이브러리 (<https://jcheminf.biomedcentral.com/articles/10.1186/s13321-021-00501-7>)를 활용해 fingerprint 형태로 변환 후 입력값으로 사용함. 각 타겟마다 독성 분율이 약 5%로 불균형 데이터이므로, random oversampling으로 보완했음. 분류모델은 Random Forest, AdaBoost, XGBoost, LightGBM을 사용했으며, 이들 중 Random Forest, LightGBM 모델의 hyperparameter tuning을 통해 최적 파라미터를 적용함.