

07 질병 연관 유전자 발굴을 위한 머신러닝 기법 설계

소속 정보컴퓨터공학부

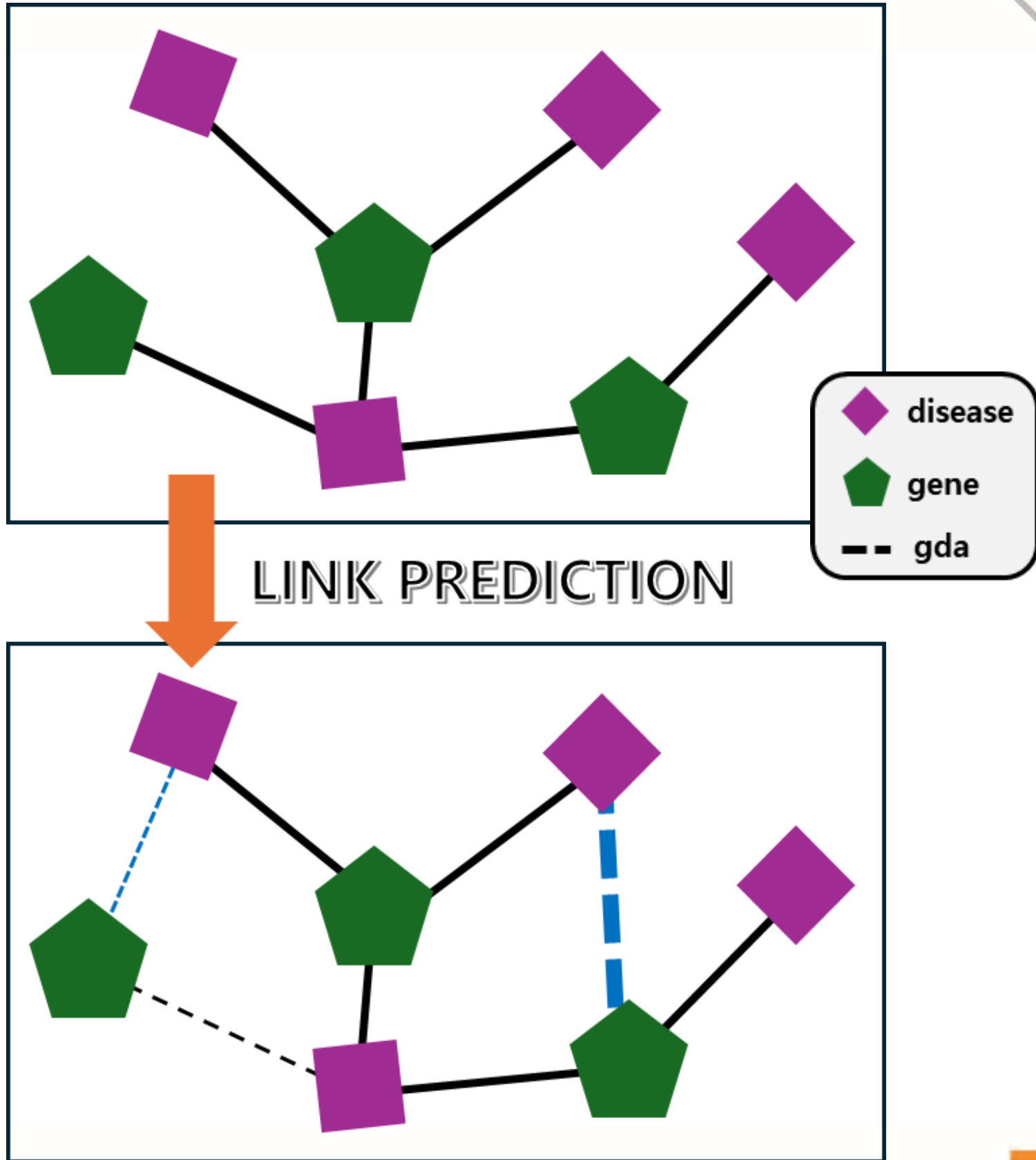
분과 A

팀명 서울대구부산

참여학생 김이경, 박화성, 이윤재

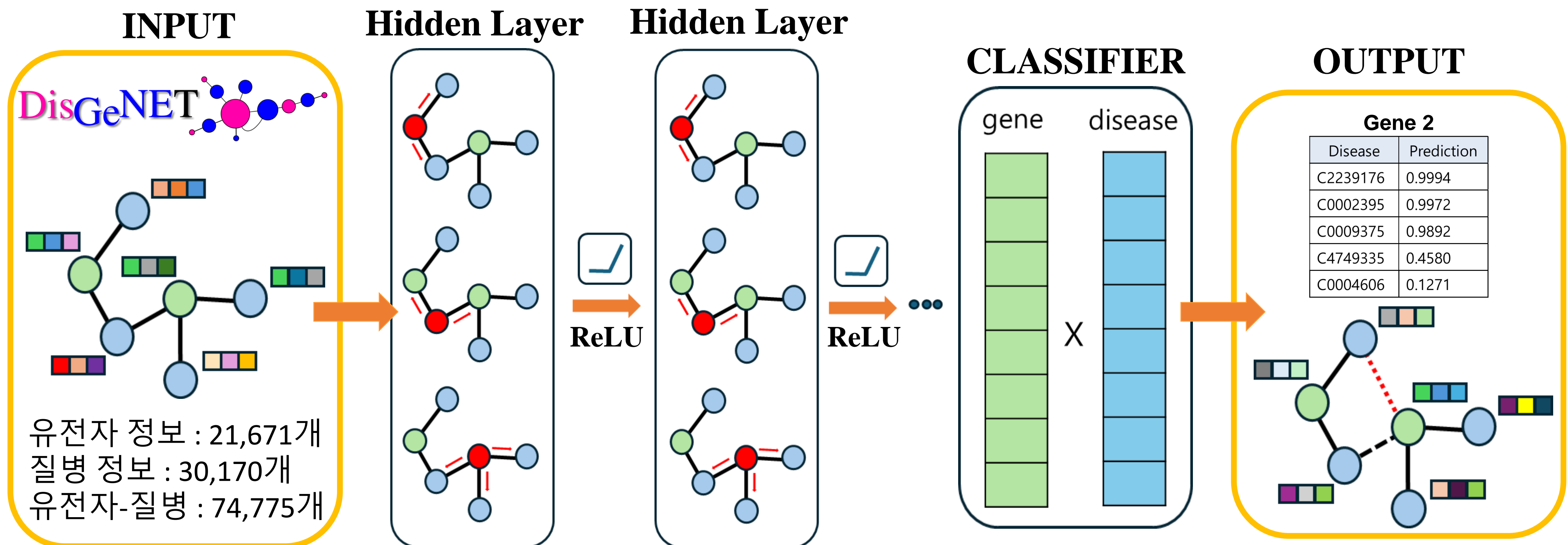
지도교수 송길태

연구 목표

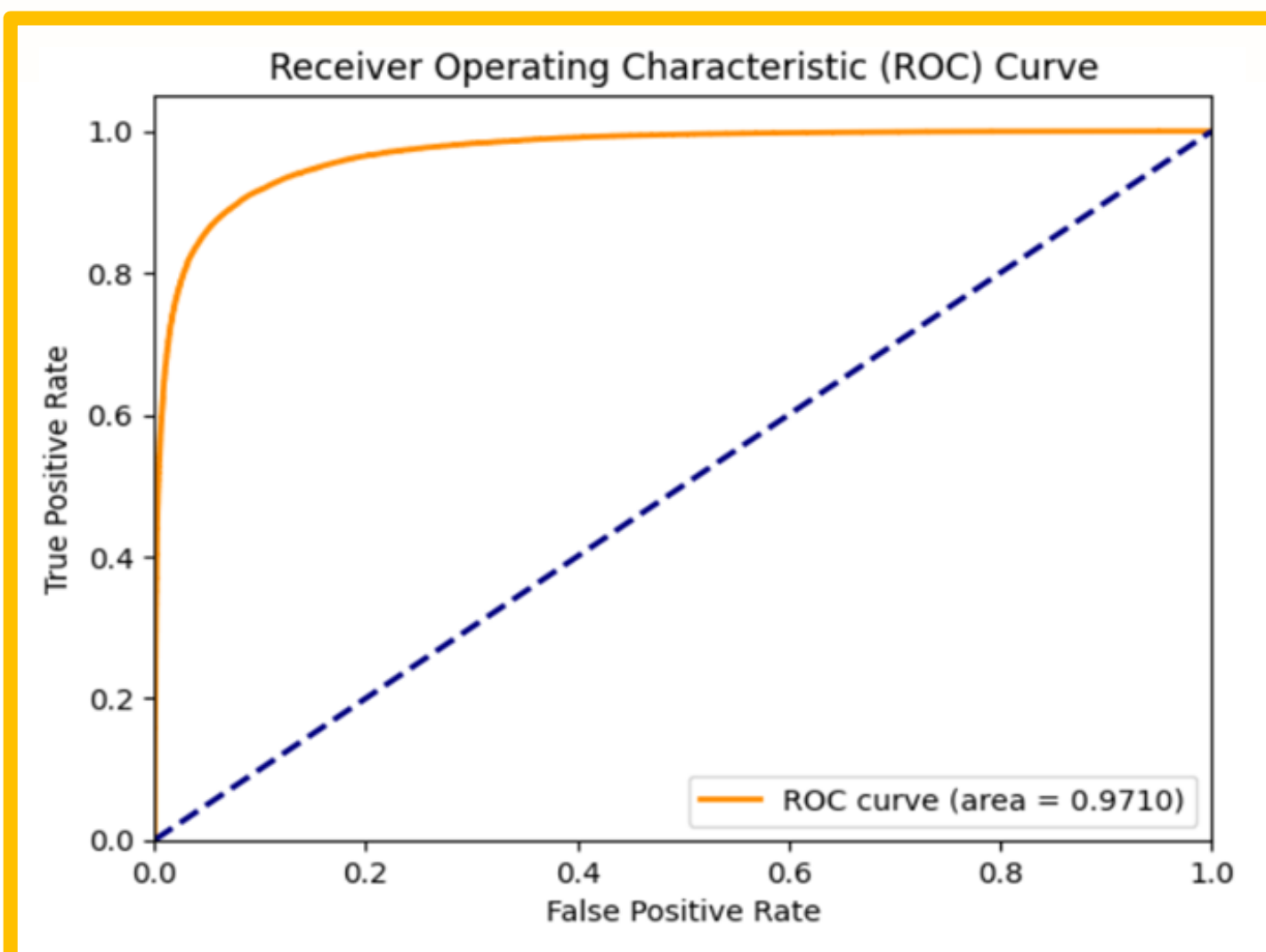


- ❖ 딥러닝 기술을 활용해 질병과 유전자 간의 연관 관계를 예측한다. 이 관계를 표현하기 위해 그래프를 구성하고, 이를 바탕으로 GCN(Graph Convolution Network) 기법을 적용한다.
- ❖ 이기종 그래프(Heterogeneous Graph) 개념을 도입하여 질병과 유전자 간의 연관 관계를 그래프로 나타낸다. 질병과 유전자는 노드, 질병과 유전자 사이의 관계는 간선으로 나타낸다.
- ❖ 상관관계가 없는 질병과 유전자를 음의 간선으로 연결하는 Negative Sampling 기법을 통해 음의 간선과 양의 간선을 분류하는 2진 분류 문제로 과제를 재정의한다.
- ❖ 시험적으로 모델에 최적화된 하이퍼파라미터를 찾아낸다.

연구 내용



최종 결과



- ❖ GCN 기술을 활용한 링크 예측(Link Prediction)을 통해 질병과 유전자 사이의 연관 관계를 예측할 수 있었다. 이를 통해 딥러닝 기법이 임상시험이나 유전자 분석 기법의 한계를 극복할 수 있을 것으로 생각한다.
- ❖ 반복적인 실험을 통해 모델에 최적화된 하이퍼파라미터를 찾아냈으며, 해당 파라미터를 바탕으로 약 **0.96**의 AUROC를 갖는 모델을 구축할 수 있었다.
- ❖ 향후 연구에서는 RNA, 단백질 등 다양한 노드를 포함하고, 질병 간 또는 유전자 간의 간선을 연결한 이기종 그래프를 구성한 후, 이를 입력으로 사용하여 노드 간의 연관 관계를 계산하는 GCN 모델을 연구하고자 한다.