

그래프 어텐션 네트워크를 이용한 항암제 조합의 시너지 효과 예측

김 상 민¹ · 유 선 용^{2*}¹전남대학교 지능전자컴퓨터공학과 석사과정²전남대학교 지능전자컴퓨터공학과 교수

Predicting Synergistic Anticancer Drug Combinations via Graph Attention Network

Sang-min Kim¹ · Sun-yong Yoo^{2*}¹Master's Course, Department of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea²Professor, Department of Intelligent Electronics and Computer Engineering, Chonnam National University, Gwangju 61186, Korea

[요 약]

항암 약물 조합의 시너지 효과 예측은 효과적인 암 치료에 필수적이다. 기존 계산적 접근법은 사전에 정의된 분자 지문에 의존하여 분자 구조를 직접 학습하지 못한다는 한계가 있다. 본 연구에서는 48종 약물과 13개 세포주로 구성된 3,014개 약물 조합 데이터를 사용하여, 그래프 어텐션 네트워크로 분자 그래프 구조를 학습하고 약물 유도 유전자 발현 정보를 통합하여 시너지 점수를 예측하였다. 모델은 MSE 63.53 ± 7.78 , 피어슨 상관계수 0.70 ± 0.04 를 달성하여 기존 방법들보다 우수한 성능을 보였다. 또한 어텐션 가중치 분석을 통해 시너지 효과에 중요한 분자 하부구조를 식별하였으며, 이는 알려진 약리학적 메커니즘과 잘 일치하였다. 이러한 결과는 신약 개발 과정에서 항암 약물 조합을 선별하는 도구로 활용될 수 있음을 시사한다.

[Abstract]

Predicting the synergistic effects of anticancer drug combinations is essential for effective cancer treatment. Existing computational approaches often rely on predefined molecular fingerprints, which may fail to identify molecular structures and atomic interactions. Utilizing a dataset of 3,014 drug combinations, encompassing 48 drugs across 13 cell lines we employed graph attention networks to extract structural features directly from chemical graphs. These structural representations were subsequently integrated with drug-induced gene expression profiles to predict synergy scores. The model achieved a mean squared error of 63.53 ± 7.78 and Pearson correlation coefficient of 0.70 ± 0.04 , outperforming existing methods. Additionally, critical molecular substructures contributing to synergistic effects were identified through attention weight analysis; these findings corresponded closely with established pharmacological mechanisms. These results suggest that our approach can serve as a robust tool for screening anticancer drug combinations within drug development.

색인어 : 약물조합, 시너지효과, 그래프 어텐션 네트워크, 약물 유도 유전자 발현, 항암치료**Keyword** : Drug Combination, Synergistic Effects, Graph Attention Networks, Drug-Induced Gene Expression, Cancer Treatment<http://dx.doi.org/10.9728/dcs.2026.27.3.831>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 05 January 2026; Revised 30 January 2026

Accepted 02 February 2026

*Corresponding Author; Sun-yong Yoo

Tel: +82-2-2861-5010

E-mail: syyoo@jnu.ac.kr

1. 서론

암은 전 세계적으로 두 번째로 많은 사망 원인으로, 매년 약 1천만 명의 사망을 초래한다[1]. 의학 연구에서 상당한 진전이 이루어졌음에도 불구하고, 암 치료 비용은 환자, 가족 및 의료 시스템에 여전히 큰 부담을 주고 있다. 따라서 효과적인 암 치료법을 찾는 것이 중요하다. 전통적인 암 치료의 단일 요법 접근법은 약물 내성, 불충분한 효능 및 중양 이질성으로 인해 한계가 있다[2]-[5]. 이러한 한계를 해결하기 위해, 단일 약제 대신 두 가지 이상의 약물을 함께 사용하는 약물 병용 요법이 최근 유망한 전략으로 부상하고 있다[6]-[8]. 병용 요법의 효과는 종종 시너지 효과에 의존하는데, 이는 두 약물의 결합 효과가 개별 효과의 합보다 클 때로 정의된다[9]. 이러한 병용 접근법은 상당한 이론적·임상적 이점을 제공한다. 여기에는 작용 기전의 상호 보완을 통한 치료 효능 향상, 다중 표적 억제로 인한 내성 발생 가능성 감소, 그리고 전체 치료 효과를 유지하거나 개선하면서 독성을 최소화하기 위해 개별 약물의 저용량 사용 가능성 등이 포함된다[10], [11]. 따라서 시너지 효과를 나타내는 약물 조합을 식별하는 것은 암 치료에서 중요한 과제이다[12].

효과적인 약물 조합을 찾는 것은 전통적으로 실험적 스크리닝에 의존해 왔다. 하지만 가능한 약물 조합의 수가 방대하기 때문에 전통적인 실험적 접근법은 시간과 노동력이 많이 들고 비용이 높아 비효율적이다[13]. 따라서 약물 조합 발견을 위한 실험적 방법을 보완하기 위해 컴퓨터 기반 접근법이 필요하다. 이러한 과제를 해결하기 위해 통계적 방법, 네트워크 기반 접근법, 기계 학습 기술 등 다양한 계산적 접근법이 개발되었다[14]-[17]. 그러나 이러한 접근법들은 복잡한 비선형 관계를 포착하고 다양한 유형의 생물학적 데이터를 효과적으로 통합하는 데 한계가 있다[18].

이러한 한계를 극복하기 위해 최근 머신러닝 기반 접근법이 약물 조합 시너지 예측에서 주목받고 있다. 연구자들은 RF(Random Forest) 및 SVM(Support Vector Machine)과 같은 머신러닝 알고리즘을 사용하여 약물 조합의 시너지 효과를 예측하였다[19], [20]. 또한 XGBoost와 그래디언트 부스팅과 같은 앙상블 방법들도 다양한 분자적 특징과 생물학적 데이터를 결합하여 높은 예측 성능을 달성하였다[21]. 이러한 모델들은 주로 분자 지문(molecular fingerprints)을 사용하여 약물의 구조적 특징을 표현한다. 분자 지문은 ECFP(Extended Connectivity Fingerprint)나 MACCS(Molecular ACCess System)와 같이 분자 구조를 고정된 길이의 이진 벡터로 변환하는 방법으로, 약물-표적 상호작용 정보나 약리학적 프로파일과 함께 머신러닝 모델의 입력으로 사용된다[22]. 최근에는 deep forest와 같은 딥러닝 방법도 제안되었다. Deep forest는 약물의 분자 지문, 물리화학적 특성(physicochemical properties), 그리고 약물 처리 후 세포주의 유전자 발현 데이터를 통합하여 시너지 효과를 예측한다[23]. 특히 약물 유도 유전자 발현 프로파일을 활용하여 약물이 세포에 미치는

전사체 수준의 변화를 반영함으로써, 소규모 데이터셋에서도 우수한 예측 성능을 달성하였다.

그러나 기존 접근법들은 사전에 정의된 분자 지문에 의존하기 때문에, 분자의 복잡한 그래프 구조와 원자 간 상호작용을 직접적으로 학습하지 못한다는 한계가 있다. 이러한 한계를 극복하기 위해 최근 그래프 신경망(GNN : graph neural network)이 약물 분자 표현 학습에 사용되고 있다. GNN은 분자를 원자(노드)와 화학 결합(엣지)으로 구성된 그래프로 표현하여, 분자의 구조와 화학적 특성을 함께 학습한다[24]. 이러한 접근법은 분자 지문과 달리 원자 간의 연결성과 화학적 환경을 직접 모델링하여 분자의 구조 정보를 효과적으로 학습할 수 있다. 대표적인 GNN 구조로는 그래프 컨볼루션 네트워크(GCN : Graph Convolutional Network)와 그래프 어텐션 네트워크(GAT : Graph Attention Network)가 있으며, 이들은 이웃 노드의 정보를 집계하여 각 노드의 표현을 학습한다[25]. 특히 GAT는 어텐션 메커니즘을 통해 이웃 원자들에 대한 가중치를 동적으로 학습하여, 시너지 효과에 중요한 원자와 결합을 선택적으로 강조할 수 있다.

본 연구에서는 GAT를 사용하여 항암제 조합의 시너지 효과를 예측하는 모델을 제안한다. GAT로 약물 분자의 그래프 구조로부터 특징을 추출하고, 약물 유도 유전자 발현 데이터를 결합하여 시너지 점수를 예측한다. 이를 통해 약물의 구조 정보와 세포의 생물학적 반응을 함께 고려한 시너지 효과 예측이 가능하다.

II. 데이터 및 방법론

2-1 실험 데이터 수집 및 전처리

본 연구에서는 이전 연구에서 사용된 약물 조합 데이터셋을 활용하였다[23]. 이 데이터셋에는 48종의 서로 다른 약물과 13개의 세포주로 구성된 총 3,014개의 약물 조합 샘플이 포함되어 있다. 약물 쌍의 시너지 효과를 정량화하기 위해 DrugComb 데이터베이스에서 개발된 S 점수 지표를 사용하였다[26]. 시너지 지표들은 약물 조합의 실제 효과와 기대되는 약물 효과를 비교하여 시너지를 평가하며, 기대 효과를 설정하는 가정에 따라 여러 참조 모델로 구분된다. HSA 모델은 두 약물 중 더 효과적인 단일 약물의 효과를 기대 효과로 사용하고, Bliss 모델은 두 약물이 독립적으로 작용한다고 가정하며, Loewe 모델은 두 약물이 동일 메커니즘으로 작용한다고 가정하고, ZIP 모델은 두 약물이 서로의 효력에 영향을 주지 않는다고 가정한다. S 점수는 교차 설계 실험에서 한 약물을 IC50 농도에 고정하고 다른 약물의 농도를 변화시키며 측정된 약물 조합 민감도(CSS: combination sensitivity score)와 단일 약물들로부터 예측된 기대 효과의 차이로 계산되며, 기존 참조 모델들과의 비교 검증을 통해 시너지 및 길항 조합을 높은 정확도로 구분할 수 있음이 확인되었다[26]. S 점수가 양수일 경우 시너지적 상호작용을, 음수일 경

우 길항적 상호작용을 의미한다. 또한 모든 약물의 화학 구조 정보는 PubChem 데이터베이스에서 표준 SMILES (simplified molecular-input line-entry system) 형식으로 수집하였다.

약물 유도 유전자 발현 프로파일은 LINCS (Library of Integrated Network-Based Cellular Signatures) L1000 데이터베이스에서 획득하였다[27]. 이 데이터베이스는 2만 개 이상의 화합물이 여러 암 세포주에 미치는 유전자 발현 변화 정보를 포함하고 있다. LINCS L1000 데이터셋은 전체 전사체의 약 80%에 해당하는 978개 주요 유전자의 발현 수준을 측정한다. 본 연구에서는 반복 실험 결과를 통합한 시그니처인 Level 5 데이터를 사용하였다. 이 데이터는 대조군 대비 발현 변화의 크기와 방향을 나타내는 z-score 형태로 제공되며, 양수는 발현 증가를, 음수는 발현 감소를 의미한다.

2-2 방법

1) 그래프 특징 추출

본 연구에서는 분자 구조를 문자열로 표현하는 SMILES 표기법을 사용하여 약물을 표현하였다. SMILES는 분자의 화학 구조를 문자열로 표현하는 표준화된 방법으로, 원자의 종류, 결합 유형, 고리 구조 등 계산 모델링에 필요한 약물의 분자적 특성과 구조 정보를 포괄적으로 포함하고 있다. SMILES 표현을 기계학습에 적합한 구조화된 데이터로 변환하기 위해 오픈소스 화학정보학 소프트웨어인 RDKit을 활용하였다. SMILES 문자열의 화학적 정확성을 검증한 후, 이를 원자를 노드로, 화학 결합을 엣지로 하는 분자 그래프 객체로 변환하였다. 이러한 분자 그래프로부터 RDKit을 사용하여 포괄적인 원자 및 결합 특징 벡터를 추출하였다. 원자 특징은 원자 기호, 차수, 전하, 혼성화, 방향족성, 수소 개수, 키랄성 등으로 구성된다. 원자 특징의 구체적인 내용은 표 1에 제시하였으며, 모든 범주형 특징은 딥러닝 모델의 입력을 위해 원-핫 인코딩 방식으로 변환하였다.

표 1. 분자 그래프 표현을 위한 원자 특징 및 차원

Table 1. Atomic features and dimensions for molecular graph representation

Atom Features	Size	Description
Atom symbol	9	[B, C, Cl, F, I, N, O, P, S]
Degree	8	Covalent bonds 0-7 (one-hot vector)
Charge	1	Formal charge
Radical electrons	1	Number of radical electrons
Hybridization	7	S, SP, SP2, SP3, SP3D, SP3D2, other
Aromaticity	1	Atom is aromatic (true/false)
Hydrogens	5	Total number of connected hydrogens (0-4)
Chirality type	2	R or S
Chirality	1	Atom chirality (true/false)

2) 모델 구조

본 연구에서는 항암제 조합의 시너지 효과를 예측하기 위한 딥러닝 모델을 제안한다. 모델은 두 약물의 분자 그래프와 각 약물을 처리했을 때의 유전자 발현 정보를 입력으로 받아 시너지 점수를 예측한다. 전체 모델은 크게 세 가지 주요 구성요소로 이루어져 있다: (1) 약물 특징 추출, (2) 유전자 발현 특징 추출, (3) 시너지 점수 예측 모듈.

각 약물의 분자 구조는 GAT를 통해 처리된다. GAT는 어텐션 메커니즘을 활용하여 분자 그래프 내에서 각 원자(노드)가 이웃 원자들에 대해 서로 다른 중요도를 학습한다. 이를 통해 약물의 화학적 구조와 결합 패턴을 포착하며, 최종적으로 분자 전체를 대표하는 하나의 고정 길이 벡터로 변환된다. 유전자 발현 데이터는 다층 퍼셉트론(MLP : Multi-Layer Perceptron)을 통해 처리된다. 각 약물의 978차원 유전자 발현 프로파일은 독립적인 MLP를 통과하며, 고차원의 유전자 발현 데이터를 저차원의 임베딩 벡터로 변환한다.

최종적으로 두 약물의 분자 임베딩 벡터와 유전자 발현 임베딩 벡터를 연결(concatenation)하여 하나의 통합 특징 벡터를 생성한다. 이 통합 벡터는 완전 연결 층(FCL : fully connected layer)을 거치며, 각 층에서는 ReLU 활성화 함수와 드롭아웃(dropout)을 적용하여 과적합을 방지하며, 최종적으로 약물 조합의 시너지 점수를 예측한다. 전체 모델 구조는 그림 1에 제시되어 있다. 본 모델에서 GAT를 통해 학습된 분자 그래프는 약물의 화학 구조를 포착하여 약물의 구조적 특징을 표현한다. 유전자 발현 정보는 약물 처리 시 세포의 전사체 수준 반응을 반영하여 약물의 생물학적 효과를 표현한다. 이 두 정보를 통합함으로써 약물의 구조와 세포 반

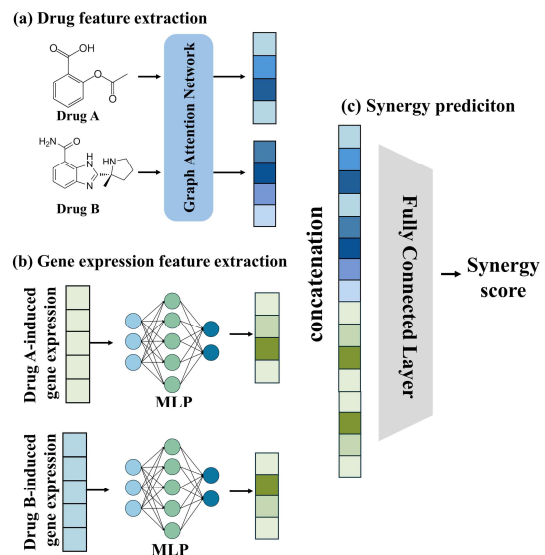


그림 1. 그래프 어텐션 네트워크를 활용한 항암제 조합 시너지 효과 예측 방법 개요

Fig. 1. Overview of method for predicting synergistic effects of anticancer drug combinations using graph attention network

응을 함께 고려하여 시너지 효과를 예측한다.

3) 그래프 어텐션 네트워크

GAT는 그래프 신경망에 어텐션 메커니즘을 도입하여 분자 그래프의 각 원자가 이웃 원자들에 서로 다른 중요도를 부여할 수 있게 한다[25]. 기존 GCN이 모든 이웃 노드를 동일하게 취급하는 것과 달리, GAT는 어텐션 스코어를 통해 각 이웃의 기여도를 동적으로 학습한다.

GAT의 핵심은 어텐션 메커니즘을 활용한 노드 특징 집계이다. 노드 i 의 특징 h_i 를 업데이트하기 위해, 먼저 이웃 노드 $j \in N(i)$ 에 대한 어텐션 계수 e_{ij} 를 계산한다. 어텐션 계수는 노드 i 와 이웃 노드 j 간의 상대적 중요도를 나타내는 값으로, 다음과 같이 계산된다.

$$e_{ij} = \text{LeakyReLU}(a^T [W h_i || W h_j]) \tag{1}$$

여기서 W 는 노드 특징의 선형 변환 행렬, a 는 학습 가능한 어텐션 가중치 벡터, $||$ 는 연결 연산을 의미한다. LeakyReLU 활성화 함수는 음수 입력에 대해서도 작은 기울기를 유지하여 그래디언트 소실 문제를 완화한다. 계산된 어텐션 계수 e_{ij} 는 소프트맥스 함수를 통해 정규화되어 최종 어텐션 스코어 α_{ij} 로 변환된다. 어텐션 스코어 α_{ij} 는 이웃 노드 j 가 중심 노드 i 의 표현을 업데이트하는 데 얼마나 기여하는지를 나타낸다.

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})} \tag{2}$$

계산된 어텐션 스코어를 사용하여 노드 i 의 새로운 표현 h'_i 를 다음과 같이 업데이트한다.

$$h'_i = \sigma \left(\sum_{j \in N(i)} \alpha_{ij} W h_j \right) \tag{3}$$

여기서 σ 는 비선형 활성화 함수이다. GAT는 멀티헤드 어텐션을 사용하여 안정성과 표현력을 향상시킨다. K 개의 어텐션 헤드를 사용할 경우, 최종 노드 표현은 다음과 같이 계산된다

$$h'_i = \parallel_{k=1}^K \sigma \left(\sum_{j \in N(i)} \alpha_{ij}^k W^k h_j \right) \tag{4}$$

분자 그래프에서 GAT는 각 원자가 주변 원자들과의 화학적 상호작용을 학습하며, 어텐션 메커니즘을 통해 중요한 결합이나 작용기에 더 높은 가중치를 부여할 수 있다. 이를 통해 약물의 분자 구조에서 시너지 효과에 중요한 구조적 특징

을 효과적으로 포착하고, 모델의 해석 가능성을 높일 수 있다. GAT를 통해 업데이트된 각 노드의 특징 벡터는 global mean pooling을 거쳐 하나의 그래프 임베딩으로 통합된다. Global mean pooling은 분자 그래프 내 모든 노드 특징의 평균을 계산하여 전체 분자를 대표하는 고정 길이의 벡터를 생성한다(그림 2). 생성된 그래프 임베딩은 약물 분자의 구조적 정보를 응축한 표현으로, 다음 단계에서 유전자 발현 임베딩과 결합되어 최종 시너지 점수 예측에 활용된다.

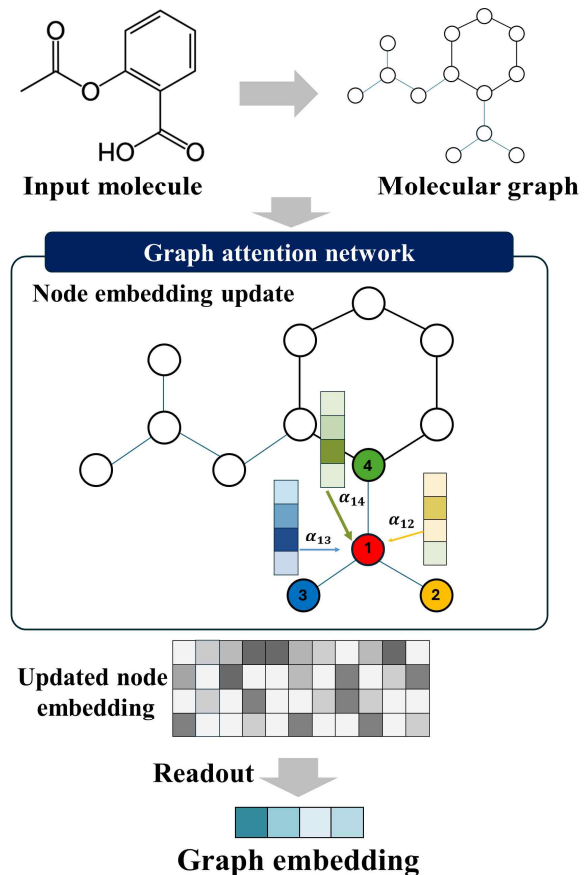


그림 2. 그래프 어텐션 네트워크의 분자 그래프 임베딩 추출 과정
Fig. 2. Molecular graph embedding extraction process in graph attention network

2-3 평가 방법

제안된 모델을 평가하기 위해 10-fold 교차검증을 수행하였다. 회귀 예측 작업의 평가 지표로 평균 제곱 오차(MSE; mean square error), 평균 제곱근 오차(RMSE; root mean square error), 평균 절대 오차(MAE; mean absolute error), 결정계수(R^2), 피어슨 상관계수(PCC; Pearson correlation coefficient)를 사용하였다. MSE는 예측값과 실제값의 차이를 제공하여 평균한 값으로, 큰 오차에 더 큰 가중치를 부여한다. RMSE는 MSE의 제곱근으로 원래 데이터와 동일한 단위를 가지며, 예측 오차의 크기를 직관적으로 이

해할 수 있게 한다. MAE는 절대 오차의 평균으로, 모든 오차를 동등하게 취급하여 이상치의 영향을 덜 받는다. R²는 모델이 데이터의 분산을 얼마나 잘 설명하는지를 나타내며, 1에 가까울수록 좋은 모델을 의미한다. PCC는 예측값과 실제값 간의 선형 상관관계를 측정하여 예측의 경향성을 평가한다. 최종 성능 점수는 모든 k-fold에 걸친 평균과 표준편차로 보고되었다.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \tag{5}$$

$$RMSE = \sqrt{MSE} \tag{6}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \tag{7}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{8}$$

$$PCC = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}} \tag{9}$$

여기서 y_i 는 실제값, \hat{y}_i 는 예측값, \bar{y} 는 실제값의 평균, n 은 샘플 수를 나타낸다.

K-fold 교차검증은 전체 데이터셋을 k개의 동일한 크기 폴드로 분할한 뒤, 하나의 폴드를 검증 데이터로, 나머지 k-1개 폴드를 훈련 데이터로 사용하는 검증 기법이다. 이러한 과정을 K번 반복 수행하여 각기 다른 데이터 분할에서의 모델 성능을 측정함으로써, 특정 데이터 분할에 의한 성능 편향을 최소화할 수 있다.

III. 실험결과

3-1 성능 평가

제안된 모델의 성능을 평가하기 위해 SVM, RF, XGBoost 등 머신러닝 방법 및 Deep Forest와 비교 평가를 수행하였다. 모델 학습에 사용된 하이퍼파라미터는 표 2에 제시하였으며, 10-fold 교차검증을 통해 얻은 5가지 회귀 지표에 대한 성능 비교 결과는 표 3에 제시되어 있다. MSE 지표에서

SVM은 107.11 ± 16.68, RF는 76.34 ± 11.35, XGBoost는 69.42 ± 8.95, Deep Forest는 64.20 ± 9.34를 달성하였으며, 본 모델은 63.53 ± 7.78로 가장 낮은 오차를 보였다. RMSE와 MAE에서도 각각 7.95 ± 0.51, 5.61 ± 0.22를 기록하여 모든 비교 모델보다 우수한 성능을 나타냈다. R² 점수는 0.49 ± 0.06으로 Deep Forest(0.47 ± 0.04)보다 높았으며, 피어슨 상관계수는 0.70 ± 0.04로 Deep Forest(0.69 ± 0.03)와 유사한 수준을 달성하였다. 이러한 결과는 GAT를 통한 분자 그래프 구조 학습과 MLP를 통한 유전자 발현 정보 처리가 약물 조합 시너지 예측에 효과적임을 보여준다.

표 2. 모델 하이퍼파라미터

Table 2. Model hyperparameters

Hyperparameters	Value
GAT hidden dimension	128
GAT attention heads	8
Embedding dimension	128
FC layer dimensions	256, 128
Dropout rate	0.2
Batch size	32
Learning rate	0.0001
Number of epochs	200
Optimizer	Adam

표 3. 모델 성능

Table 3. Model performance

Method	MSE (± std)	RMSE (± std)	MAE (± std)	R2 (± std)	PCC (± std)
SVM	107.11 ± 16.68	10.32 ± 0.82	6.96 ± 0.45	0.13 ± 0.03	0.43 ± 0.06
Random Forest	76.34 ± 11.35	8.71 ± 0.69	5.96 ± 0.37	0.38 ± 0.05	0.63 ± 0.05
XGBoost	69.42 ± 8.95	8.31 ± 0.56	5.81 ± 0.33	0.44 ± 0.05	0.66 ± 0.04
Deep forest	64.20 ± 9.34	7.99 ± 0.58	5.64 ± 0.23	0.47 ± 0.04	0.69 ± 0.03
Ours	63.53 ± 7.78	7.95 ± 0.51	5.61 ± 0.22	0.49 ± 0.06	0.70 ± 0.04

3-2 모델 해석가능성 분석

모델의 해석가능성을 평가하기 위해 GAT의 어텐션 가중치를 분석하여 시너지 효과에 중요한 분자 하부구조를 식별하였다. 모델이 예측한 상위 시너지 점수를 기준으로 5-FU(5-Fluorouracil)와 ABT-888(Veliparib) 조합을 선정하여 분석을 수행하였다. ABT-888은 PARP(Poly ADP-ribose polymerase) 억제제로, DNA 손상 복구 경로를 차단한다. 어텐션 분석 결과, 아마이드 그룹(amide group)에서 높은 어텐션 가중치가 관찰되었으며(그림 3), 이는 PARP 활성 부위에 결합하는 핵심 구조로 알려져 있다 [28]. 5-FU의 경우 플루오린(F)과 카보닐기(C=O)에서 높은

가중치를 나타냈다. 5-FU는 우라실 유사체로 세포 내에서 FdUMP로 대사되며, thymidylate synthase를 비가역적으로 억제하여 DNA 합성에 필수적인 thymidine 생성을 차단한다[29].

이러한 구조적 특징은 두 약물의 시너지 메커니즘과 연결된다. 5-FU에 의한 DNA 합성 억제는 DNA 손상을 유발하고, ABT-888은 PARP 매개 복구 경로를 차단하여 손상된 DNA의 축적을 증가시킴으로써 암세포 사멸을 유도한다 [30]. 어텐션 분석을 통해 식별된 핵심 작용기들은 이러한 약리학적 메커니즘과 일치하며, GAT가 약물의 구조-활성 관계를 학습하고 시너지 효과에 중요한 분자 구조를 효과적으로 포착할 수 있음을 보여준다.

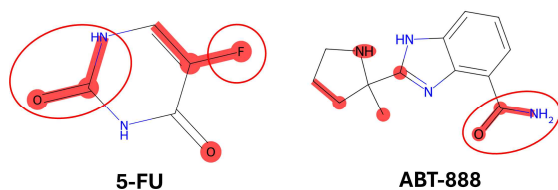


그림 3. 어텐션 가중치 기반 분자 하부구조 분석 결과
Fig. 3. Attention weight-based molecular substructure analysis

IV. 결 론

항암 약물 조합의 시너지 효과 예측은 효과적인 암 치료를 위해 중요한 역할을 한다. 이를 위해 다양한 머신러닝 및 딥러닝 방법들이 활용되고 있지만, 기존 접근법들은 사전에 정의된 분자 지문에 의존하기 때문에 분자의 복잡한 그래프 구조와 원자 간 상호작용을 직접적으로 학습하지 못한다는 한계가 있다.

본 연구에서는 GAT를 사용하여 분자의 그래프 구조를 학습하고, 약물 유도 유전자 발현 데이터를 함께 처리하여 약물 조합의 시너지 점수를 예측하는 모델을 제안하였다. 제안된 모델은 10-fold 교차검증에서 MSE 63.53 ± 7.78 , RMSE 7.95 ± 0.51 , MAE 5.61 ± 0.22 , R^2 0.49 ± 0.06 , 피어슨 상관계수 0.70 ± 0.04 를 달성하여, SVM, Random Forest, XGBoost 등 기존 머신러닝 방법보다 우수한 성능을 보였다. 특히 비교 모델 중 가장 우수한 성능을 보인 Deep Forest보다 높은 성능을 보여주었다.

본 연구에서 제안된 모델은 신약 개발 초기 단계에서 대규모 약물 조합 스크리닝이 필요한 상황에 실질적으로 활용될 수 있다. 실험 검증 전에 계산적 예측을 통해 시너지 효과가 높은 조합을 우선 선별함으로써 시간과 비용을 크게 절감할 수 있으며, 어텐션 가중치 분석을 통해 약물의 핵심 작용기를 식별함으로써 시너지 메커니즘에 대한 이해를 높일 수 있다. 다만 본 연구는 48종의 약물과 13개 세포주로 구성된 데이터

셋을 사용하였으므로, 결과의 일반화 가능성에는 제한이 있다. 특히 제한된 수의 세포주로 인해 특정 암종이나 세포에 대한 편향이 존재할 가능성이 있다. 향후 연구에서는 더 많은 약물과 다양한 암세포주를 포함한 대규모 데이터셋으로 모델을 확장하여 예측 성능과 범용성을 검증할 필요가 있다. 또한 FDA 승인 약물 데이터베이스나 임상 시험 데이터를 활용하여 실제 임상 환경에서의 적용 가능성을 평가하는 것이 필요하다. 이를 통해 본 연구에서 제안한 접근법은 보다 안전하고 효과적인 항암 약물 조합을 발굴하는 데 중요한 도구로 활용될 수 있을 것이다.

감사의 글

본 연구는 보건복지부의 재원으로 한국보건산업진흥원을 통해 수행된 연구(RS-2025-19252970), 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(RS-2025-16063391), 식품의약품안전처의 연구개발비(RS-2025-02215961, RS-2024-00332003)로 이루어진 연구로서, 관계부처에 감사드립니다.

참고문헌

- [1] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R. L. Siegel, I. Soerjomataram, and A. Jemal, "Global Cancer Statistics 2022: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries," *CA: A Cancer Journal for Clinicians*, Vol. 74, No. 3, pp. 229-263, May 2024. <https://doi.org/10.3322/caac.21834>
- [2] N. Vasan, J. Baselga, and D.M. Hyman, "A View on Drug Resistance in Cancer," *Nature*, Vol. 575, 7782, pp. 299-309, November 2019. <https://doi.org/10.1038/s41586-019-1730-1>
- [3] C. Holohan, S. Van Schaeybroeck, D. B. Longley, and P. G. Johnston, "Cancer Drug Resistance: An Evolving Paradigm," *Nature Reviews Cancer*, Vol. 13, No. 10, pp. 714-726, 2013. <https://doi.org/10.1038/nrc3599>
- [4] M. G. Rees, L. Brenan, M. do Carmo, P. Duggan, B. Bajrami, M. Arciprete, ... and C. M. Johannessen, "Systematic Identification of Biomarker-Driven Drug Combinations to Overcome Resistance," *Nature Chemical Biology*, Vol. 18, No. 6, pp. 615-624, 2022. <https://doi.org/10.1038/s41589-022-00996-7>
- [5] A. Zafar, S. Khatoun, M. J. Khan, J. Abu, and A. Naeem, "Advancements and Limitations in Traditional Anti-Cancer Therapies: A Comprehensive Review of Surgery, Chemotherapy, Radiation Therapy, and Hormonal Therapy,"

- Discover Oncology*, Vol. 16, No. 1, 607, 2025. <https://doi.org/10.1007/s12672-025-02198-8>
- [6] B. Al-Lazikani, U. Banerji, and P. Workman, "Combinatorial Drug Therapy for Cancer in the Post-Genomic Era," *Nature Biotechnology*, Vol. 30, No. 7, pp. 679-692, July 2012. <https://doi.org/10.1038/nbt.2284>
- [7] R. Bayat Mokhtari, T. S. Homayouni, N. Baluch, E. Morgatskaya, S. Kumar, B. Das, and H. Yeger, "Combination Therapy in Combating Cancer," *Oncotarget*, Vol. 8, No. 23, pp. 38022-38043, June 2017. <https://doi.org/10.18632/oncotarget.16723>
- [8] J. Lehár, A. S. Krueger, W. Avery, A. M. Heilbut, L. M. Johansen, E. R. Price, ... and A. A. Borisy, "Synergistic Drug Combinations Tend to Improve Therapeutically Relevant Selectivity," *Nature Biotechnology*, Vol. 27, No. 7, pp. 659-666, July 2009. <https://doi.org/10.1038/nbt.1549>
- [9] T.-C. Chou, "Drug Combination Studies and Their Synergy Quantification Using the Chou-Talalay Method," *Cancer Research*, Vol. 70, No. 2, pp. 440-446, January 2010. <https://doi.org/10.1158/0008-5472.CAN-09-1947>
- [10] A. C. Palmer and P. K. Sorger, "Combination Cancer Therapy Can Confer Benefit via Patient-to-Patient Variability without Drug Additivity or Synergy," *Cell*, Vol. 171, No. 7, pp. 1678-1691, e13, December 2017. <https://doi.org/10.1016/j.cell.2017.11.009>
- [11] D. A. Yardley, "Drug Resistance and the Role of Combination Chemotherapy in Improving Patient Outcomes," *International Journal of Breast Cancer*, Vol. 2013, 137414, June 2013. <https://doi.org/10.1155/2013/137414>
- [12] T. Pemovska, M. Kontro, B. Yadav, H. Edgren, S. Eldfors, A. Szwarzda, ... and K. Wennerberg, "Individualized Systems Medicine Strategy to Tailor Treatments for Patients With Chemorefractory Acute Myeloid Leukemia," *Cancer Discovery*, Vol. 3, No. 12, pp. 1416-1429, December 2013. <https://doi.org/10.1158/2159-8290.CD-13-0350>
- [13] R. Lin and G. Yin, "Bayesian Optimal Interval Design for Dose Finding in Drug-Combination Trials," *Statistical Methods in Medical Research*, Vol. 26, No. 5, pp. 2155-2167, 2017. <https://doi.org/10.1177/0962280215594494>
- [14] J. O'Neil, Y. Benita, I. Feldman, M. Chenard, B. Roberts, Y. Liu, ... and S. D. Shumway, "An Unbiased Oncology Compound Screen to Identify Novel Combination Strategies," *Molecular Cancer Therapeutics*, Vol. 15, No. 6, pp. 1155-1162, June 2016. <https://doi.org/10.1158/1535-7163.MCT-15-0843>
- [15] S. L. Holbeck, R. Camalier, J. A. Crowell, J. P. Govindharajulu, M. Hollingshead, L. W. Anderson, ... and J. H. Doroshow, "The National Cancer Institute ALMANAC: A Comprehensive Screening Resource for the Detection of Anticancer Drug Pairs With Enhanced Therapeutic Activity," *Cancer Research*, Vol. 77, No. 13, pp. 3564-3576, July 2017. <https://doi.org/10.1158/0008-5472.CAN-17-0489>
- [16] F. Cheng, I.A. Kovács, and A.-L. Barabási, "Network-Based Prediction of Drug Combinations," *Nature Communications*, Vol. 10, No. 1, 1197, March 2019. <https://doi.org/10.1038/s41467-019-09186-x>
- [17] W. Zhao, K. Sachsenmeier, L. Zhang, E. Sult, R. E. Hollingsworth, and H. Yang, "A New Bliss Independence Model to Analyze Drug Combination Data," *Journal of Biomolecular Screening*, Vol. 19, No. 5, pp. 817-821, June 2014. <https://doi.org/10.1177/1087057114521867>
- [18] X.-M. Zhao, M. Iskar, G. Zeller, M. Kuhn, V. van Noort, and P. Bork, "Prediction of Drug Combinations by Integrating Molecular and Pharmacological Data," *PLoS Computational Biology*, Vol. 7, No. 12, e1002323, December 2011. <https://doi.org/10.1371/journal.pcbi.1002323>
- [19] X. Li, Y. Xu, H. Cui, T. Huang, D. Wang, B. Lian, ... and L. Xie, "Prediction of Synergistic Anti-Cancer Drug Combinations Based on Drug Target Network and Drug Induced Gene Expression Profiles," *Artificial Intelligence in Medicine*, Vol. 83, pp. 35-43, November 2017. <https://doi.org/10.1016/j.artmed.2017.05.008>
- [20] P. Ding, R. Yin, J. Luo, and C.-K. Kwoh, "Ensemble Prediction of Synergistic Drug Combinations Incorporating Biological, Chemical, Pharmacological, and Network Knowledge," *IEEE Journal of Biomedical and Health Informatics*, Vol. 23, No. 3, pp. 1336-1345, May 2019. <https://doi.org/10.1109/JBHI.2018.2852274>
- [21] P. Sidorov, S. Naulaerts, J. Arieu-Bonnet, E. Pasquier, and P. J. Ballester, "Predicting Synergism of Cancer Drug Combinations Using NCI-ALMANAC Data," *Frontiers in Chemistry*, Vol. 7, 509, July 2019. <https://doi.org/10.3389/fchem.2019.00509>
- [22] Q. Liu and L. Xie, "TranSynergy: Mechanism-Driven Interpretable Deep Neural Network for the Synergistic Prediction and Pathway Deconvolution of Drug Combinations," *PLoS Computational Biology*, Vol. 17, No. 2, e1008653, February 2021. <https://doi.org/10.1371/journal.pcbi.1008653>
- [23] W. Lin, L. Wu, Y. Zhang, Y. Wen, B. Yan, C. Dai, ... and X. Bo, "An Enhanced Cascade-Based Deep Forest Model

for Drug Combination Prediction,” *Briefings in Bioinformatics*, Vol. 23, No. 2, bbab562, March 2022. <https://doi.org/10.1093/bib/bbab562>

- [24] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, “The Graph Neural Network Model,” *IEEE Transactions on Neural Networks*, Vol. 20, No. 1, pp. 61-80, January 2009. <https://doi.org/10.1109/TNN.2008.2005605>
- [25] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, “Graph Attention Networks,” arXiv:1710.10903, 2018. <https://doi.org/10.48550/arXiv.1710.10903>
- [26] B. Zagidullin, J. Aldahdooh, S. Zheng, W. Wang, Y. Wang, J. Saad, ... and J. Tang, “DrugComb: An Integrative Cancer Drug Combination Data Portal,” *Nucleic Acids Research*, Vol. 47, No. W1, pp. W43-W51, July 2019. <https://doi.org/10.1093/nar/gkz337>
- [27] A. Subramanian, R. Narayan, S. M. Corsello, D. D. Peck, T. E. Natoli, X. Lu, ... and T. R. Golub, “A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles,” *Cell*, Vol. 171, No. 6, pp. 1437-1452, e17, November 2017. <https://doi.org/10.1016/j.cell.2017.10.049>
- [28] U. K. Velagapudi, É. Rouleau-Turcotte, R. Billur, X. Shao, M. Patil, B. E. Black, ... and T. T. Talele, “Novel Modifications of PARP Inhibitor Veliparib Increase PARP1 Binding to DNA Breaks,” *Biochemical Journal*, Vol. 481, No. 6, pp. 437-460, March 2024. <https://doi.org/10.1042/BCJ20230406>
- [29] D. B. Longley, D. P. Harkin, and P. G. Johnston, “5-Fluorouracil: Mechanisms of Action and Clinical Strategies,” *Nature Reviews Cancer*, Vol. 3, No. 5, pp. 330-338, May 2003. <https://doi.org/10.1038/nrc1074>
- [30] S. Paul, S. Chatterjee, S. Sinha, S. R. Dash, R. Pradhan, B. Das, ... and C. N. Kundu, “Veliparib (ABT-888), A PARP Inhibitor Potentiates the Cytotoxic Activity of 5-Fluorouracil by Inhibiting MMR Pathway Through Dereglulation of MSH6 in Colorectal Cancer Stem Cells,” *Expert Opinion on Therapeutic Targets*, Vol. 27, No. 10, pp. 999-1015, 2023. <https://doi.org/10.1080/14728222.2023.2266572>

김상민(Sang-min Kim)



2024년~현 재: 전남대학교 지능전자컴퓨터공학과 석사과정
※ 관심분야: 생명정보학(bioinformatics), 인공지능(artificial intelligence)

유선용(Sun-yong Yoo)



2012년: 한국항공대학교
정보통신공학과 (공학석사)
2018년: 한국과학기술원
바이오및뇌공학과 (공학박사)

2018년~2019년: 국민건강보험공단 빅데이터실 부연구위원
2019년~현 재: 전남대학교 지능전자컴퓨터공학과 교수
※ 관심분야: 생명정보학(bioinformatics), 인공지능(artificial intelligence), 빅데이터(big data) 등