

# 유전자 Attention 잔차 학습 딥 러닝 모델을 통한 mRNA와 miRNA 발현 데이터 기반의 Low-Grade Glioma 환자의 생존분석 모델

최상혁, 구혜은, 이민혁\*  
중앙대학교 전자전기공학부

choiroy@cau.ac.kr, hae10202@cau.ac.kr, \*mlee@cau.ac.kr

## A Deep Learning Model with Gene Attention and Residual Learning using mRNA and miRNA Expression Data for Survival Estimation of Low-Grade Glioma

Choi Sanghyuk Roy, Ku Hyeen, Lee Minhyeok\*  
Chung-Ang University.

### 요 약

딥 러닝 모델과 유전자 발현 데이터를 이용한 Low-Grade Glioma (LGG) 생존 분석은 최근 몇 년 동안 집중적으로 연구되고 있다. 그러나 기존 연구는 다른 발현 데이터와 임상 데이터를 무시하고 mRNA 발현 데이터만 고려한 경우가 대부분이었다. 본 연구에서는 이러한 한계를 해결하기 위해 mRNA와 miRNA를 동시에 사용하는 딥 러닝 모델을 제안한다. 제안한 모델을 사용하여 MYBL1과 hsa-mir-421은 각각 LGG 환자의 예후에서 가장 유의한 mRNA와 miRNA로 확인되었다. LGG 환자의 생존에 영향을 미치는 mRNA와 miRNA를 추정하는 다수의 기존 연구는 본 연구의 결과를 뒷받침한다.

### I. 서 론

본 논문에서는 mRNA 데이터뿐만 아니라 miRNA 데이터를 동시에 활용하는 딥 러닝 모델을 제안한다. 이 연구의 동기는 멀티-오믹스 체계를 가진 이러한 관점이 LGG 환자의 생존 분석에 대한 더 나은 이해를 제공할 수 있다는 생각에서 비롯된다. miRNA와 관련된 추가 정보를 고려할 때, 확인된 예후 유전자는 mRNA만을 사용할 때 확인된 유전자와 구별될 수 있으며, 환자의 예후 측면에서 mRNA와 miRNA 사이의 관계를 설명할 수 있다. mRNA와 miRNA 사이의 높은 상관관계를 활용하는 여러 연구가 있지만, 본 연구에서는 두 타입의 데이터를 동시에 사용 가능한 유전자 어텐션 레이어를 제안한다. 또한 miRNA 데이터를 입력으로 사용하는 유전자 어텐션 레이어를 통해 어떤 miRNA가 LGG 환자의 생존과 가장 관련성이 높은지 추정할 수 있다.

### II. 본론

어텐션 메커니즘은 딥 러닝 레이어의 출력 벡터에 0에서 1까지의 특정 값을 곱하는 방법이다. 이 방법은 예측 대상과 관련이 적은 형상의 효과를 감소시킨다. 따라서 값이 0일수록 해당 형상이 대상과 관계를 가질 확률이 낮다는 것을 나타낸다. 예를 들어, LGG 환자의 생존 일수가 딥 러닝 모델의 예측 대상으로 학습될 경우, 0에 가까운 어텐션 값은 입력된 해당 유전자가 LGG 환자의 생존과 관련이 적다는 것을 의미한다. 이러한 방법을 통해 LGG의 생존과 관련이 높은 유전자를 선별할 수 있다. 이처럼, 이 프로세스는 덜

관련된 기능을 제거하기 위한 필터로 간주될 수 있다. 이 메커니즘은 다음과 같은 방정식으로 나타낼 수 있다.

$$X_{k+1} := A^T \odot \tau_k(W_k \cdot X_k)$$

이 때,  $X_k$ 는  $k$  번째 레이어의 입력,  $W_k$ 는  $k$  번째 레이어의 파라미터,  $\tau_k$ 는  $k$  번째 레이어의 비선형 활성화함수,  $A^T$ 는 어텐션 값이다.

$A^T$ 를 얻기 위해서는 다음과 같은 또 다른 딥 러닝 레이어가 사용된다:

$$A := \sigma(W_{A2} \cdot \text{ReLU}(W_{A1} \cdot X_0)).$$

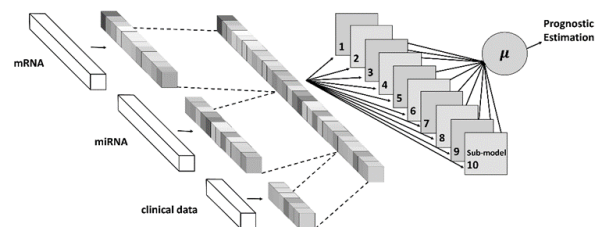


그림 1 제안하는 모델의 구조

본 연구에서는 여러 모듈을 가진 딥 러닝 앙상블 모델을 제시한다. 앙상블 딥 러닝 모델의 각 모듈은 동일한 구조를 사용한다. 가중치 매개 변수와 미니 배치 샘플링의 무작위 초기화 때문에 동일한 구조가 사용되더라도 이러한 모듈은 서로 다른 결과를 생성한다. 이러한 다양한 결과는 결과가 다양한 기존의 딥 러닝 훈련의 무작위성을 상쇄할 수 있다. 앙상블 딥 러닝 모델에서 여러 모듈의 결과가 평균화되므로 이 문제를 완화할 수 있다.

본 연구에서는 모델을 평가하기 위해 TCGA-LGG(Low-Grade Glioma) 데이터 세트를 사용하였다. 이 실험에서는 125 개의 중도절단되지 않은 샘플이 5 겹 교차 검증을 수행하기 위해 사용되었다. 샘플 중 100 개의 샘플을 무작위로 선택하여 모델을 훈련시키고 나머지 샘플은 각 교차 검증에서 테스트 세트가 된다. 이 5 겹 교차 검증은 20 회 수행되므로 서로 다른 훈련과 테스트 세트를 사용하여 100 개의 실험을 수행하였다. 전처리 과정의 경우, 유전자의 0 값이 샘플의 절반을 초과하면, 유전자는 데이터 세트에서 누락된다. 또한 표본에 결측값이 있는 경우 해당 유전자는 제외되었다. 유전자 발현 데이터는 로그 정규화에 의해 사전 처리되었다.

표 1 LGG 환자의 생존 추정 결과 C-Index

Proposed	Ridge	SVM	RF	Coxnet
0.70	0.69	0.70	0.70	0.68

제안하는 모델, Ridge regression, SurvivalSVM, RSF, Coxnet 의 C-Index 로 측정된 성능 비교는 표 1 와 같다. C-Index 는 20 번의 5 겹 교차 검증을 통해 100 번 평가되었다. 그 결과 제안하는 모델의 C-Index 는 0.70 이었다. 제안하는 모델은 은 Ridge 와 Coxnet 보다 좋은 성능을 보여주었고, 그 두 모델의 성능은 각각 0.69 과 0.68 이었다. 요약하자면, 제안하는 모델은 다른 기존 방법과 경쟁적인 성능을 보여주었다.

본 연구에서 제시된 방법으로 추정된 LGG 환자의 생존과 관련된 mRNA 와 miRNA 는 다음과 같다:

표 2 LGG 환자의 생존 관련 mRNA 와 miRNA

순위	mRNA	miRNA
1	MYBL	mir-421
2	CARHSP1	mir-885
3	C9orf50	mir-495

MYBL1 은 유전자 어텐션 값이 약 0.88 로 가장 유의한 mRNA 로 추정되었다. 이는 MYBL1 이 LGG 의 예후와 직접적인 관계를 가질 가능성이 있음을 나타낸다. 이러한 제안된 모델의 추정은 MYBL1 이 종양 발생과 높은 관련이 있는 세포의 생존, 증식 및 분화를 조절한다고 제안한 여러 기존 연구와 일치한다 [1].

또한, 예후 miRNA 는 제안된 모델에 의해 확인되었으며, 이는 제안된 모델의 유전자 어텐션 계층에서 비롯되는 mRNA 및 기타 miRNA 와의 관계를 고려했다. 그 결과, hsa-mir-421 은 주의 값이 0.60 으로 가장 유의한 miRNA 로 추정되었으며, 다른 추정된 예후 miRNA 는 hsa-mir-885, hsa-mir-495 였다. 이러한 결과는 위암 [2,3], 신경아세포종 [4], 폐선암 [5], 유방암[6], 골육종[7]을 포함한 다양한 암에서 hsa-mir-421 의 과발현을 입증하는 기존 연구의 결과와 일치한다.

### III. 결론

본 논문에서는 LGG 와 예후 유전자의 예후를 추정하기 위해 멀티-오믹스 체계를 갖춘 딥 러닝 모델을 제안하였다. Multi-PEN 의 구조는 miRNA 발현과 임상 데이터를 위한 어텐션 계층이 제안된 구조에 사용되었다. 유전자 어텐션 메커니즘의 이점 때문에, 제안된 구조는

생존 관련 mRNA 와 miRNA 를 식별하는 데 사용될 수 있다.

### ACKNOWLEDGMENT

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1F1A1050977).

### 참 고 문 헌

- [1] Cicirò, Y.; Sala, A. "MYB oncoproteins: emerging players and potential therapeutic targets in human cancer." *Oncogenesis* 2021, 10, 1-15.
- [2] Wu, J. et al. "MicroRNA-421 is a new potential diagnosis biomarker with higher sensitivity and specificity than carcinoembryonic antigen and cancer antigen 125 in gastric cancer." *Biomarkers* 2015, 20, 58-63.
- [3] Liu, H. et al. "Correlation between microRNA-421 expression level and prognosis of gastric cancer." *International Journal of Clinical Experimental Pathology* 2015, 8, 15128.
- [4] Li, Y. et al. "Downregulation of tumor suppressor menin by miR-421 promotes proliferation and migration of neuroblastoma." *Tumor Biology* 2014, 35, 10011-10017.
- [5] Cineaglia, N.C. et al. "Integrative transcriptome analysis identifies deregulated microRNA-transcription factor networks in lung adenocarcinoma." *Oncotarget* 2016, 7, 28920.
- [6] Lou, W. et al. "Five miRNAs-mediated PIEZO2 downregulation, accompanied with activation of Hedgehog signaling pathway, predicts poor prognosis of breast cancer." *Aging* 2019, 11, 2628.
- [7] Zhou, S. et al. "miR-421 is a diagnostic and prognostic marker in patients with osteosarcoma." *Tumor Biology* 2016, 37, 9001-9007.