

딥러닝을 활용한 오토인코더 기반 질병 관련 miRNA 예측 연구

하 지 환*

Autoencoder-based Disease-related miRNA Prediction Research using Deep Learning

Ji-Hwan Ha*

이 논문(또는 저서)은 부경대학교 자율창의학술연구비(2021년)에 의하여 연구되었음.

요 약

MicroRNA(miRNA)는 단일염기 가닥의 non coding small RNA로서, 유전자 발현 제어를 통해 인간의 다양한 질병에서 중요한 역할을 하는 조절 인자로 알려져 있다. 따라서, 질병 관련 miRNA를 추론하는 것은 인간의 복잡한 질병 메커니즘을 이해하는데 있어서 중요한 밑거름이 되고 있다. 이러한 관점에서 miRNA와 질병 간의 관계를 추론하기 위한 다양한 인공지능 기반 모델들이 제안되어왔다. 본 논문에서는 질병 관련 miRNA를 효과적으로 예측하기 위한 딥러닝 기반 모델을 제안한다. 오토인코더를 활용하여 miRNA와 질병 특징 벡터를 효과적으로 추출하였고, 이를 딥러닝 모델의 입력값으로 사용하여 질병 관련 miRNA 예측 모델을 보다 정확하게 훈련하였다. Leave-One-Out Cross-Validation(LOOCV) 검증 방법을 활용하여 모델의 예측 성능을 정확하게 평가하였으며, 결과적으로 제안하는 모델의 성능이 AUC(0.9075) 측면에서 기존 모델들보다 더 나은 성능을 도출하는 것을 확인할 수 있었다.

Abstract

MicroRNA(miRNA)s are small non-coding RNAs that play significant roles in various human complex diseases. Therefore, the identification of disease-related miRNAs have become a crucial task to understand the human disease mechanisms at the molecular level. At this regard, various artificial intelligence-based models have been developed to predict the relationships between miRNAs and diseases. In this paper, we propose novel deep learning-based computational model to effectively predict the disease-related miRNAs. First, we mapped miRNA and disease feature vectors by utilizing auto-encoder. Then, we used each miRNA and disease feature vector as input for the deep learning architecture model. Finally, we performed the leave-one-out cross-validation(LOOCV) to validate the prediction performance of our model. The experimental results demonstrated that superior performance of our model by comparing the area under the receiver operating characteristic curve(AUC) scores with other three existing methods.

Keywords

microRNA, deep learning, machine learning, biomarker

* 부경대학교 데이터정보과학부
빅데이터융합전공 교수
- ORCID: <https://orcid.org/0000-0002-6086-5693>

· Received: Mar. 10, 2022, Revised: Apr. 08, 2022, Accepted: Apr. 11, 2022
· Corresponding Author: Ji-Hwan Ha
Division of Data Information Science, Major of Big Data Convergence,
Pukyong National University, Busan 48513, Korea
Tel.: +82-51-629-4614, Email: jhha@pknu.ac.kr

1. 서 론

MicroRNA(miRNA)는 19~25 Nucleotide 길이의 Non coding small RNA로서, DNA 유전정보로부터 단백질이 합성되는 과정에서 mRNA의 3'-untranslated region(3'-UTR)에 불완전한 상보결합을 통해 유전자 발현을 억제하는 것으로 알려져 있다 [1][2]. 또한, 수많은 연구 결과를 통해 세포 분화 (Cell differentiation), 성장(Growth), 면역반응(Immune reaction) 등 여러 생물학적 메커니즘에서 중요한 역할을 하는 것으로 밝혀졌으며, 질병 발병에도 큰 영향을 미친다는 것이 수많은 생물연구를 통해 밝혀지고 있다[3]-[7].

또한, 인간 생체 내에서 발생하는 생물학적 메커니즘은 복잡한 인과관계에서 발생되기 때문에, 질병 발병 인자로서 miRNA의 역할을 분석하는 것은 생물정보학 연구 측면에서 매우 중요한 과제이다. 이로 인해 miRNA와 질병 간의 관계를 추론하는 연구는 인간 질병 메커니즘의 이해를 위한 발판을 마련할 뿐만 아니라, 새로운 바이오마커 추출 관점에서 큰 생물학적 의미를 지니고 있다.

최근 인공지능 기술과 데이터마이닝 기술이 분야를 막론하고 막강한 성능을 도출함에 따라, 많은 연구에서 인공지능 모델을 활용하여 질병과 miRNA의 관계를 추론하는 연구가 진행되어왔다[8]-[14]. Chen et al.은 miRNA 간의 유사성 정보를 기반으로 miRNA 네트워크를 구축하였다. 구축된 miRNA 네트워크를 기반으로 랜덤 워크(Random walk) 알고리즘을 수행하여 질병과 miRNA의 관계를 추론하였다 [15]. 또한, Chen et al.은 앙상블 기법 기반 의사 결정 나무(Decision tree) 알고리즘을 활용하여 새로운 예측 모델을 제안하였다. 제안하는 모델에서 PCA 기법을 활용하여 차원 축소를 수행하였고, 이를 통해 질병 관련 miRNA를 효과적으로 추출하였다[16].

Shi et al.은 miRNA 타겟 정보와 PPI(단백질-단백질 상호작용) 네트워크를 활용하여 특정 질병과 miRNA 간의 관계를 예측 및 분석하였다[17]. PPI 네트워크란 단백질들 사이의 관계에 대한 통합 시스템 및 생물학 네트워크로서 세포 내에서 DNA 복제 및 전사, 다양한 신호전달 경로 및 대사를 매개

하는 기능적인 핵심 단위이다. 제안하는 모델에서는 PPI 네트워크를 활용하여 랜덤 워크를 수행하였으며, 단백질 정보를 활용하여 질병과 miRNA의 관계를 추론하였다. Mørk et al. 단백질과 miRNA 간의 관계 데이터, 질병과 단백질 간의 관계 데이터를 활용하여 질병 관련 miRNA 추출하였다.

즉, 특정 miRNA와 단백질 관계가 존재하고 해당 단백질과 질병 간의 관계가 존재한다면, 특정 miRNA는 질병과 관련이 높을 것이라는 가정에 모델을 설계하였다[18]. Ha et al.은 추천시스템에서 널리 쓰이는 인공지능 기법인 행렬 분해 알고리즘(Matrix factorization)을 적용하여 질병 관련 miRNA 추출 모델을 설계하였다[19]. 행렬 분해 알고리즘은 관측데이터의 특징(Feature)들을 Latent space라 불리는 임베딩 공간에 개체들의 특징들을 매핑(Mapping)시키는 기법으로서 추천시스템에서 사용자(User)와 아이템(Item) 간의 관계를 추론할 때 널리 사용되는 인공지능 기법이다. 저자들은 인공지능 기반 행렬 분해 알고리즘을 활용하여 miRNA와 질병 간의 관계를 효과적으로 예측하는 모델을 제안하였다. 하지만 기존 연구에서의 공통적인 한계점은 miRNA와 질병을 정확하게 표현할 수 있는 특징 벡터를 효과적으로 추출하지 못한다는 것이며, 이는 곧 예측 성능을 저하시킨다는 문제로 이어지고 있다[20].

본 논문에서는 딥러닝을 모델에 활용하여 효과적으로 질병과 miRNA의 연관성을 추론하고자 한다. 오토인코더(Auto-encoder)를 통해 miRNA와 질병 특징 벡터(Feature vector)를 효과적으로 추출하여 수치형 벡터로 표현하였다. 이를 딥러닝 모델 훈련(Training)을 위한 입력 값으로 사용하였고, 효과적으로 miRNA와 질병 간의 관계를 추론하였다.

본 논문의 구성은 다음과 같다. 1장에서는 질병 관련 miRNA 예측 연구에 대한 배경과 관련 연구를 소개하고, 2장 본론에서는 제안하는 모델에서 사용되는 데이터와 방법론에 대해 자세히 기술한다. 3장 실험 부분에서는 다양한 실험을 통해 본 논문에서 제안하는 방법론의 성능을 분석하고 실험 결과를 통해 제안하는 모델의 우수성을 증명한다. 4장 결론 부분에서는 추후 향후 계획과 한계점에 대해 언급한다.

II. 딥러닝 기반 질병 관련 예측 모델링

2.1 데이터

많은 연구를 통해 질병 발병 인자로서의 miRNA 기능들이 다양한 생물학적 근거로 입증되고 있다. 이로 인해, 다양한 온라인 데이터베이스에서 질병과 miRNA의 관계 정보를 제공하고 있다. 본 논문에서는 질병 관련 miRNA 정보를 제공하는 HMDD v3.2 온라인 데이터베이스에서 질병과 miRNA 관계 정보를 각각 다운로드 받아 모델의 훈련 및 검증 데이터로 사용하였다[21].

본 논문에서는 검증된 관계 값을 사용하기 위해, 생물학 실험을 통해 증명된 데이터베이스를 사용하였다. HMDD v3.2는 1,206개의 miRNA와 893개의 질병 사이에서 실제 문헌 정보에서 관계가 증명된 35,547개의 샘플 데이터를 제공하고 있다.

2.2 오토인코더 기반 특징 추출

본 연구에서는 딥러닝을 기반으로 하여 miRNA와 질병 간의 관계를 효과적으로 예측하는 모델을 제안하고자 한다. 질병과 miRNA의 관계를 효과적으로 추출하기 위해서는 miRNA, 질병 특징 벡터를 각각 정확하게 수치형 데이터로 표현하는 기법이 중요하다. 본 논문에서는 딥러닝 모델의 입력값으로 사용하기 위한 miRNA, 질병 특징 벡터를 인공신경망 기반 오토인코더를 활용하여 차원 축소를 통해 추출하였다. 오토인코더란 인공신경망을 활용한 비지도 학습 방법(Unsupervised learning) 기반 표현 학습(Representation learning) 기법이다.

이는 입력 데이터값의 차원을 축소하여 원래의 데이터 값보다 작은 차원의 잠재 공간(Latent space)로 압축시킨 후, 이를 다시 원본 입력으로 복원할 수 있도록 네트워크를 학습시키는 방법이다. 저차원으로 압축시키는 부분을 인코더(Encoder)라 하며, 압축된 데이터를 원래의 차원으로 복원시키는 부분을 디코더(Decoder)라 일컫는다(그림 1).

인코더 부분에서는 원-핫 인코딩(One-hot encoding)을 통해 표현된 고차원(1,206) miRNA m 벡터를 인코딩(Encoding) 과정을 통해 5차원의 h 벡터로 매핑

시킨다(식 (1)).

$$h_m = s(Wm + b) \tag{1}$$

s 는 비선형(Non-linear)함수인 시그모이드(Sigmoid) 함수로 표현되며, W 는 입력 데이터값을 은닉층(Hidden layer)으로 표현되는 저차원 임베딩 공간으로 연결해주는 가중치 값들이다. 디코딩 단계에서는 저차원으로 표현된 h_m 벡터를 복원과정을 통해 다시 miRNA m' 벡터로 변환한다(식 (2)).

$$m' = s(W'h_m + b') \tag{2}$$

복원된 m' 는 새롭게 예측된 miRNA 벡터이며, 데이터 샘플을 통해 손실함수의 에러 값을 줄여나가는 방법으로 W, W', b, b' 로 표현된 오토인코더의 가중치 값들을 수정한다. 손실함수 $L(m, m')$ 는 식 (3)과 같이 표현된다.

$$L(m, m') = - \sum_{k=1}^d [v_k \log v'_k + (1 - v_k) \log (1 - v'_k)] \tag{3}$$

이와 같은 방식으로 원-핫 인코딩으로 표현된 893차원 질병 벡터 d 를 5차원 벡터 h_d 로 변환하였다.

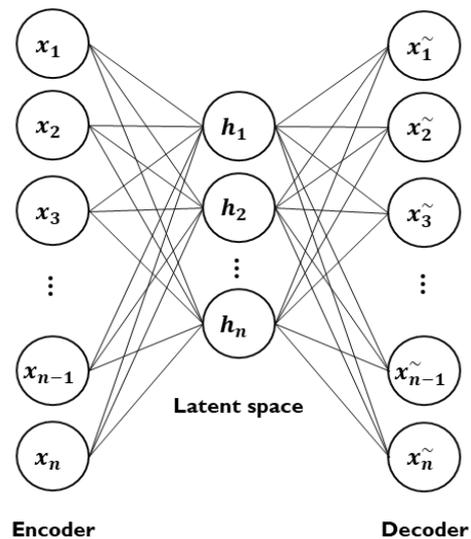


그림 1. 오토인코더 구조
Fig. 1. Structure of auto-encoder

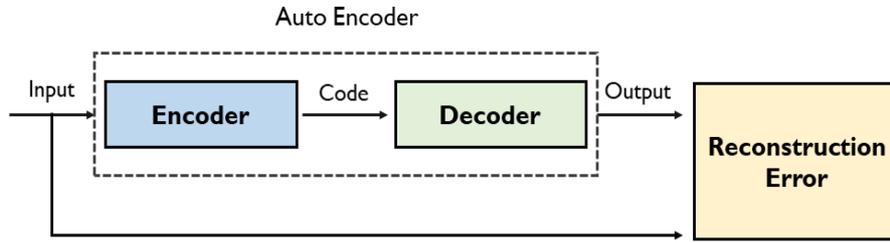


그림 2. 오토인코더 시각화
Fig. 2. Visualization of auto-encoder

오토인코더를 통해 변환된 miRNA 벡터 h_m 와 질병 벡터 h_d 를 결합(Concatenate)하여 $h_i = [h_m; h_d]$ 를 생성하였다. 최종적으로 h_i 와 질병 관계 유무를 나타내는 이진(Binary) 정보를 결합하여 딥러닝 모델의 입력 데이터값으로 사용하였다.

2.3 질병 관련 miRNA 추출 모델링

오토인코더를 통해 추출된 miRNA와 질병 벡터 정보를 결합하여 h_i 벡터를 생성하였고, 이를 딥러닝의 입력값으로 사용하였다(그림 3). 단순 결합된 벡터들은 miRNA와 질병 사이의 관계를 충분히 표현하지 못하는 경향을 지닌다. 따라서, 여러 개의 퍼셉트론 뉴런을 여러 층으로 쌓은 다층신경망 구조(MLP, Multi-Layer Perceptron)를 사용하여 예측 모델을 설계하였다(식 (4)).

다섯 개의 층을 쌓아서 다층신경망을 구성하였고, 훈련 방법으로는 드롭아웃(Dropout)을 사용하였다. 또한, 활성화 함수(Activation function)로는 ReLU를 사용하였다.

$$z_1 = h_i \tag{4}$$

$$f_1(z_1) = a_1(W_1^T z_1 + b_1)$$

...

$$f_L(z_{L-1}) = a_L(W_L^T z_{L-1} + b_L)$$

$$y_{md} = (h^T f_L(z_{L-1}))$$

a_x, b_x 그리고 w_x 는 각각 x번째 층 퍼셉트론의 활성화 함수, 바이어스(Bias) 벡터, 가중치 행렬로 표현된다.

III. 실험 결과

3.1 교차 검증 실행

본 논문에서 제안하는 모델의 예측 성능을 평가하기 위해 LOOCV(Leave-One-Out Cross Validation) 교차 검증을 수행하였다. LOOCV는 n개의 데이터 샘플이 존재할 때, 하나의 데이터 샘플을 선택하여 이를 검증 데이터로 사용하고, 나머지 n-1개의 데이터 샘플을 훈련데이터로 사용하여, 이를 n회 반복하는 성능평가 기법이다. 즉, n개의 miRNA-질병 관계 데이터가 주어졌을 때, n-1개의 miRNA-질병 관계 데이터 샘플을 통해 예측 모델을 훈련하였고, 나머지 하나의 데이터 샘플을 통해 모델의 성능을 측정하였다. 최종적으로, 이를 n회 반복 실험하여 모델의 성능을 엄격하게 평가하였다.

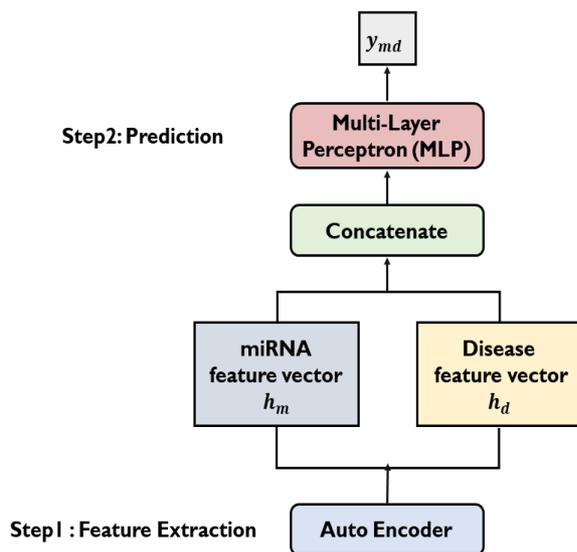


그림 3. 제안하는 모델의 전체 흐름도
Fig. 3. Workflow of the proposed model

3.2 비교 실험 실행

본 논문에서는 LOOCV 검증 방법을 사용하여, 기존 인공지능 기반 예측 모델들과의 성능평가를 실행하였다[12][19]. 질병 관련 miRNA를 정확하게 예측하였는지 평가하기 위해, AUC(Area Under the receiver operating Characteristic Curve) 값을 계산하여 기존 예측 모델들과의 비교 실험을 진행하였다. ROC(Receiver Operating Characteristic)는 분류 모델의 성능을 나타내는 그래프로써, AUC 값은 ROC 곡선 아래 영역을 의미한다. 우수한 분류 모델일수록 AUC 값이 1에 가까우며, 0.5는 모델의 분류 능력이 거의 없음을 의미한다.

그림 4에서 확인할 수 있듯이 본 논문에서 제시하는 모델이 AUC = 0.9075 값을 도출하며, 기존 방법론 MDHGI (0.9040), PMAMCA (0.8967) 보다 우수한 성능을 도출한 것을 확인할 수 있었다.

또한, 본 논문에서 제안하는 예측 모델의 성능을 확인하기 위해 추가적인 평가 지표를 적용하여 비교 실험을 진행하였다(표 1). ROC 곡선은 Positive class와 Negative class의 샘플 수가 다를 경우, 이를 정확하게 반영하지 못한다는 단점을 지니고 있다. 이에, AUPR을 적용하여 추가적인 성능 평가실험을 진행하였다. AUPR은 정밀도(Precision)-재현율(Recall) 곡선의 아래 면적을 의미한다.

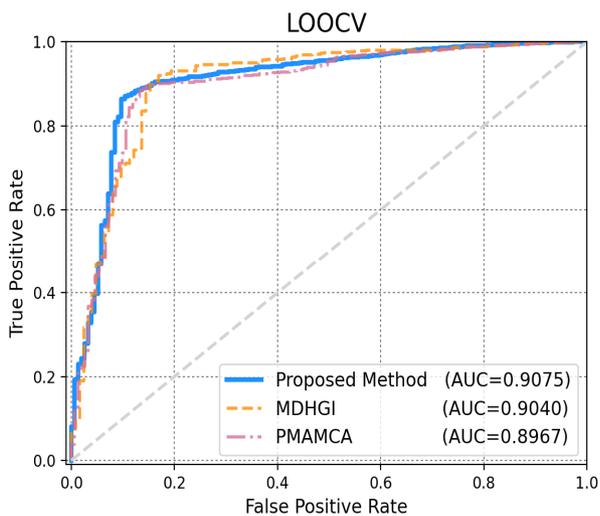


그림 4. ROC 곡선 기반 성능 비교 평가
Fig. 4. ROC curve based performance comparison evaluation

표 1. 비교 실험 결과

Table 1. Comparative experiments results

Methods	AUC	AUPR	F1
Proposed method	0.9075	0.8426	0.8814
MDHGI	0.9040	0.8404	0.8745
PMAMCA	0.8967	0.8501	0.8802

또한, F1 스코어를 통해서 본 논문에서 제안하는 모델의 성능을 평가하였다. F1 스코어는 정밀도와 재현율을 결합한 지표로서, 이는 식 (5)를 통해서 계산할 수 있다.

$$F1 = 2 * \frac{\text{정밀도} * \text{재현율}}{\text{정밀도} + \text{재현율}} \quad (5)$$

또한, 유방암 관련 miRNA 추출 비교 실험을 진행하였다. 각 논문에서 제안하는 서로 다른 방법론을 통해 유방암 관련 miRNA를 100개, 50개 추출하였을 때, 실제로 얼마나 많은 유방암 관련 miRNA를 추출하였는지 확인하였다. 그림 5에서 확인할 수 있듯이, 본 논문에서 제안하는 모델이 각각 96, 49개의 유방암 관련 miRNA를 추출하였으며, 이는 비교 모델(MDHGI: 91개, 46개, PMAMCA: 88개, 44개)보다 나은 성능임을 확인할 수 있었다.

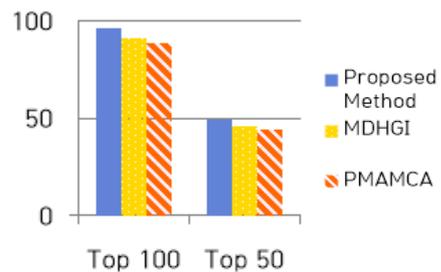


그림 5. 유방암 관련 miRNA 추출 성능 비교 평가
Fig. 5. Performance evaluation on breast cancer-related miRNA extraction

3.3 생존 분석(Survival analysis)

본 연구에서 제안하는 예측 모델의 우수성을 증명하기 위해 추가적으로 생존 분석 실험을 진행하였다. 생존 분석이란, 어떠한 사건이 발생할 때까지

의 시간을 분석하여, 두 집단의 생존율을 비교하고, 생존율에 미치는 위험인자를 분석하는 방법론이다. 주어진 시간이 지났을 때 생존하는 인구 및 환자의 비율의 문제는 생존 분석에서 가장 많이 다루는 문제이다. 즉, 주어진 인자가 중요한 역할을 하는 인자일수록, 두 집단의 생존율은 큰 차이를 보이게 된다. 생존 분석을 그림으로 도식화한 것이 Kaplan-meier plot이다. Kaplan-meier plot의 x축은 생존 기간 y축은 생존율을 각각 의미하며, 이를 통해 집단 간의 시간에 따른 생존율을 쉽게 비교할 수 있다. 본 논문에서 제안하는 방법론으로 예측한 유방암 관련 miRNA (hsa-miR-101, has-miR-210, has-miR-148a)에 대하여 유방암 환자의 생존율을 분석하였다. 특정 miRNA가 유방암과 관련이 높을수록, 두 집단의 생존율은 큰 차이를 보이게 되며, 이는 곧 모델이 예측한 miRNA가 질병 발병 인자라는 것을 뜻하게 된다.

결과적으로 그림 6, 7, 8을 통해 확인할 수 있듯이, miRNA 발현량 값이 높은 환자 집단과 발현량 값이 적은 환자 집단이 생존율 측면에서 큰 차이를 보이는 것을 확인할 수 있었으며, 이는 곧 해당 miRNA가 유방암 발병에 중요한 인자라는 것을 뜻하게 된다. 이는 통계적 수치값(p-value<0.005)으로도 실험의 타당성을 증명할 수 있었다.

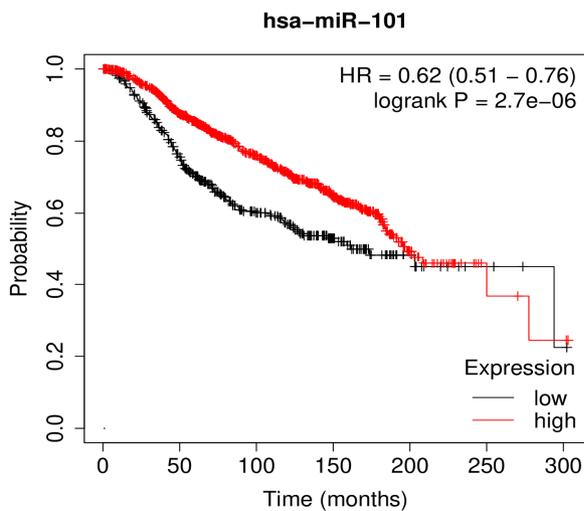


그림 6. 유방암 환자의 생존을 위한 hsa-miR-101의 Kaplan-Meier 플롯

Fig. 6. Kaplan-meier plots of hsa-miR-101 for survival of patients with breast cancer

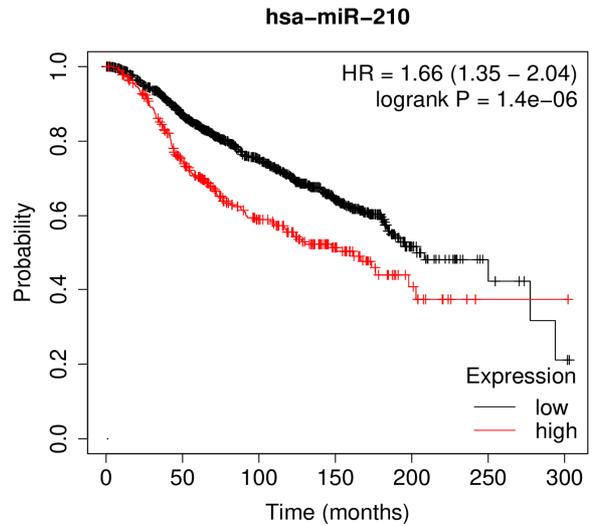


그림 7. 유방암 환자의 생존을 위한 hsa-miR-210의 Kaplan-meier 플롯

Fig. 7. Kaplan-meier plots of hsa-miR-210 for survival of patients with breast cancer

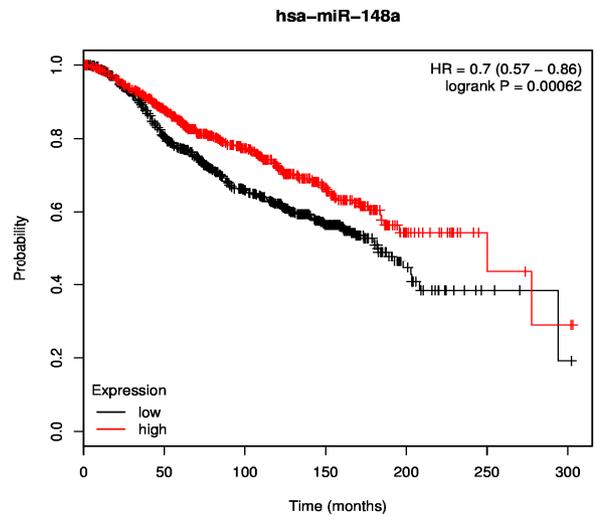


그림 8. 유방암 환자의 생존을 위한 hsa-miR-148a의 Kaplan-Meier 플롯

Fig. 8. Kaplan-meier plots of hsa-miR-148a for survival of patients with breast cancer

IV. 결론 및 향후 과제

miRNA는 번역을 하지 않는(Non-coding) 작은 RNA의 한 종류로서, 다양한 질병에서 중요한 역할을 하는 것으로 알려져 있다. 이러한 이유로 생물정보학에서는 질병과 miRNA 사이의 관계를 보다 효과적으로 추출하기 위해 많은 전산(Computational) 모델들이 제안되어왔다.

본 논문에서는 딥러닝 기반 질병 관련 miRNA 예측 모델을 제안하였다. 제안하는 딥러닝 기반 예측 모델을 정확하게 훈련하기 위해, 오토인코더를 활용하여 miRNA와 질병 특징 벡터를 각각 효과적으로 추출하였다. 또한, 추출된 특징 벡터들을 결합하여 이를 딥러닝 모델의 입력값으로 사용하였으며, 이를 통해 miRNA와 질병 사이의 관계를 효과적으로 예측하였다. 다양한 평가 지표를 기반으로 기존 모델들과의 비교 실험을 진행하였으며, 본 논문에서 제안하는 예측 모델의 우수한 성능을 확인할 수 있었다. 기계학습 기반 질병 관련 miRNA 추출 모델을 활용하여 생물학적 실험을 통해 드는 시간과 비용을 단축할 수 있었고, 실질적으로 의생명과학 분야에서의 임상 응용 가능성도 확인할 수 있었다.

본 연구에서는 오토인코더를 통해 miRNA와 질병 특징 벡터를 추출하였다. 최근 연구 결과에 따르면, 정교한 기술을 바탕으로 다양한 인공지능 기반 특징 추출 모델들이 제시되고 있다. 예측 모델의 정확성을 높이기 위해서는 개체들을 수치형 데이터로 정확하게 표현할 수 있는 벡터 표현 기법이 중요하다. 추후 연구에서는 보다, 더 정교한 인공지능 기법을 활용하여 miRNA 특징 벡터를 추출하고, 모델의 예측 성능을 높이고자 한다. 또한, miRNA 연구 특성상 이미 밝혀진 정보, 데이터가 부족하다는 특징을 지니고 있다. 인공지능 모델을 효과적으로 훈련하기 위해서는 데이터 확보가 중요시된다. 추후 연구에서는 텍스트 마이닝, 오버샘플링(Over sampling) 기법을 활용하여 충분한 데이터를 확보하고, 이를 통해 예측 모델의 성능을 높이는 발판을 마련하고자 한다.

References

- [1] V. Ambros, "The functions of animal microRNAs", *Nature*, Vol. 431, No. 7006, pp. 350-355, Sep. 2004. <https://doi.org/10.1038/nature02871>.
- [2] D. P. Bartel, "MicroRNAs: Genomics, biogenesis, mechanism, and function", *Cell*, Vol. 116, No. 2, pp. 281-297, Jan. 2004. [https://doi.org/10.1016/s0092-8674\(04\)00045-5](https://doi.org/10.1016/s0092-8674(04)00045-5).
- [3] D. P. Bartel, "MicroRNAs: Target recognition and regulatory functions", *Cell*, Vol. 136, No. 2, pp. 215-233, Jan. 2009. <https://doi.org/10.1016/j.cell.2009.01.002>.
- [4] P. Xu, M. Guo, and B. A. Hay, "MicroRNAs and the regulation of cell death", *Trends Genet.*, Vol. 20, No. 12, pp. 617-624, Dec. 2004. <https://doi.org/10.1016/j.tig.2004.09.010>.
- [5] X. Karp and V. Ambros, "Encountering microRNAs in cell fate signaling", *Science*, Vol. 310, No. 5752, pp. 1288-1289, Nov. 2005. <https://doi.org/10.1126/science.1121566>.
- [6] E. A. Miska, "How microRNAs control cell division, differentiation and death", *Current Opinion Genet. Develop.*, Vol. 15, No. 5, pp. 563-568, Oct. 2005. <https://doi.org/10.1016/j.gde.2005.08.005>.
- [7] A. M. Cheng, "Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis", *Nucleic Acids Res.*, Vol. 33, No. 4, pp. 1290-1297, Feb. 2005. <https://doi.org/10.1093/nar/gki200>.
- [8] J. Ha, C. Park, C. Park, and S. Park, "IMIPMF: Inferring miRNA-disease interactions using probabilistic matrix factorization", *Journal of Biomedical Informatics*, Vol. 102, pp. 103358, Feb. 2020. <https://doi.org/10.1016/j.jbi.2019.103358>.
- [9] J. Ha, C. Park, C. Park, and S. Park, "Improved prediction of miRNA-disease associations based on matrix completion with network regularization", *Cells*, Vol. 9, No. 4, pp. 881, Apr. 2020. <https://doi.org/10.3390/cells9040881>.
- [10] Y. Qu, H. Zhang, C. Liang, and X. Dong, "KATZMDA: Prediction of miRNA-disease associations based on KATZ model", *IEEE Access*, Vol. 6, pp. 3943-3950, Sep. 2017. <https://doi.org/10.1109/ACCESS.2017.2754409>.
- [11] X. Chen, Q. F. Wu, and G. Y. Yan, "RKNNMDA: Ranking-based KNN for miRNA-disease association prediction", *RNA Biol.*, Vol.

- 14, No. 7, pp. 952-962, Jul. 2017. <https://doi.org/10.1080/15476286.2017.1312226>.
- [12] X. Chen, J. Yin, J. Qu, and L. Huang, "MDHGI: Matrix decomposition and heterogeneous graph inference for miRNA-disease association prediction", *PLOS Comput. Biol.*, Vol. 14, No. 8, pp. 1-24, Aug. 2018. <https://doi.org/10.1371/journal.pcbi.1006418>.
- [13] X. Che, and L. Huang, "LRSSLMDA: Laplacian regularized sparse subspace learning for MiRNA-disease association prediction", *PLOS Comput. Biol.*, Vol. 13, No. 12, e1005912, Dec. 2017. <https://doi.org/10.1371/journal.pcbi.1005912>.
- [14] J. Ha, H. Kim, Y. Yoon, and S. Park, "A method of extracting disease-related microRNAs through the propagation algorithm using the environmental factor based global miRNA network", *Bio-Medical Materials and Engineering*, Vol. 26, No. s1, pp. 1763-1772, Aug. 2015. <https://doi.org/10.3233/BME-151477>.
- [15] X. Chen, M. X. Liu, and G. Y. Yan, "RWRMDA: predicting novel human microRNA-disease associations", *Molecular BioSystems*, Vol. 8, No. 10, pp. 2792-2798, Oct. 2012. <https://doi.org/10.1039/c2mb25180a>.
- [16] X. Chen, C. C. Zhu, and J. Yin, "Ensemble of decision tree reveals potential miRNA-disease associations", *PLoS Comput. Biol.*, Vol. 15, No. 7, e1007209, Jul. 2019. <https://doi.org/10.1371/journal.pcbi.1007209>.
- [17] H. Shi, J. Xu, G. Zhang, L. Xu, C. Li, L. Wang, Z. Zhao, W. Jiang, Z. Guo, and X. Li, "Walking the interactome to identify human miRNA-disease associations through the functional link between miRNA targets and disease genes", *BMC Systems Biology*, Vol. 7, No. 1, pp 1-12, Oct. 2013. <https://doi.org/10.1186/1752-0509-7-101>.
- [18] S. Mørk, S. Pletscher-Frankild, A. Palleja Caro, J. Gorodkin, and L. J. Jensen, "Protein-driven inference of miRNA-disease associations", *Bioinformatics*, Vol. 30, No. 3, pp. 392-397, Nov. 2013. <https://doi.org/10.1093/bioinformatics/btt677>.
- [19] J. Ha, C. Park, and S. Park, "PMAMCA: prediction of microRNA-disease association utilizing a matrix completion approach", *BMC Systems Biology*, Vol. 13, No. 1, pp. 1-13, Mar. 2019. <https://doi.org/10.1186/s12918-019-0700-4>.
- [20] J. Ha and C. Park, "MLMD: Metric Learning for Predicting MiRNA-Disease Associations", *IEEE Access*, Vol. 9, pp. 78847-78858, May 2021. <https://doi.org/1109/ACCESS.2021.3084148>.
- [21] Z. Huang, J. Shi, Y. Gao, C. Cui, S. Zhang, J. Li, Y. Zhou, and Q. Cui, "HMDD v3.0: a database for experimentally supported human microRNA-disease associations", *Nucleic Acids Research*, Vol. 47, No. D1, pp. 1013-1017, Oct. 2018. <https://doi.org/10.1093/nar/gky1010>.

저자소개

하 지 환 (Ji-Hwan Ha)



2015년 8월 : 연세대학교
컴퓨터과학과 (공학석사)
2020년 8월 : 연세대학교
컴퓨터과학과 (공학박사)
2020년 ~ 2021년 : 하와이 암센터
포스닥 연구원
2021년 9월 ~ 현재 : 부경대학교

데이터정보과학부 빅데이터융합전공 조교수
관심분야 : 기계학습, 생물정보학, 데이터마이닝, 딥러닝