

딥러닝을 활용한 안정형 염색체 이상 분석을 위한 염색체 검출 알고리즘

한중훈¹, 강승영¹, 이인경², 주해미², 양세정^{1*}

¹연세대학교 의공학과

²한국수력원자력(주) 방사선보건원

Deep-learning based Chromosome Detection Algorithm for Stable Chromosome Aberration

Junghun Han¹, Seungyoung Kang¹, Inkyung Lee², Haemi Joo², Sejung Yang^{1*}

¹Department of Biomedical Engineering, Yonsei University, Wonju, 26493, Republic of Korea

²Radiation Health Institute, Korea Hydro & Nuclear Power, Seoul, 04505, Republic of Korea
*syang@yonsei.ac.kr

Abstract

Chromosome aberration analysis is a method used in the process of screening or diagnosing various genetic diseases. Chromosome aberration analysis is also used in workplaces dealing with radiation materials or in estimating radiation exposure in the event of an exposure accident. However, since most chromosome aberration analysis is done manually by experts, which is time consuming and expensive. In particular, stable chromosome aberration analysis requires a fluorescence in situ hybridization process, which makes analysis more difficult. In this study, therefore, we present a chromosome detection algorithm for stable chromosome aberration analysis using deep learning. As a result of the study, the effect of detecting chromosomes through deep learning could be confirmed.

1. 연구 배경

염색체 이상 분석은 여러 유전적 질환을 선별하거나 진단하는 과정에서 사용되는 방법이다[1]. 또한 방사선 물질을 취급하는 작업장이나 방사선 노출 사고 시, 노출된 방사선량을 추정하는 방법으로 염색체 이상 분석이 활용된다. 방사선 노출 선량은 물리학적 선량 계측기를 통해 검출할 수 있지만, 실제 작업환경에서 적용하기 어렵다는 단점이 존재한다. 따라서 염색체 이상 분석으로 이상 염색체의 비율을 계산하는 생물학적 선량평가를 주로 활용한다. 하지만, 대부분의 염색체 이상 분석은 전문가에 의해 수작업으로 수행되기 때문에 많은 시간과 비용이 필요하다.

따라서, 최근에는 딥러닝을 활용하여 이러한 염색체 이상 분석을 자동화하는 연구가 활발히 진행되고 있다. 하지만, 활발한 연구에도 불구하고, 염색체 이상 분석은 특정 핵형에만 국한되거나[2], 불안정형 염색체 이상 분석을 위한 연구에만 집중하고 있다[3].

염색체 이상 분석은 불안정형 이상 염색체와 안정형 이상 염색체를 바이오 마커로 이용하여 수행되는 방법이다. 불안정형 이상 염색체는 세포 분열 시 염색체 일부가 잘려서 다른 염색체와 결합하는 이동원체성 염색체를 의미하고, 안정형 이상 염색체는 한 염색체의 일부가 다른 염색체의 일부와 전환되는 전좌가 발생한 염색체를 의미한다. 특히, 안정형 이상 염색체의 경우 세포가 분열하더라도 모양의 변화가 없기 때문에 추가적인 염색을 통해 특정 염색체를 구분해야 하는 문제가 발생한다.

안정형 염색체 이상 분석은 FISH(fluorescence in situ hybridization)를 통해 1, 2, 4번 염색체를 다른 색으로 염색하여 전좌를 구분한다. 하지만, 추가적인 염색과정에서 염색체 영상의 화질이 떨어지고, 특정 염색체의 색의 강도만 강해지는 문제가 발생한다. 이러한 문제들로 인해, 안정형 염색체 이상 분석 과정은 현재까지 전문가에 의해 수작업으로 진행되고 있는 상황이다.

본 연구에서는 안정형 염색체 이상 분석을 위해 딥러닝을 활용하여 FISH 염색된 영상에서 염색체를 검출하는 알고리즘을 제시한다. 제시하는 알고리즘은 FISH 염색 과정에서 발생하는 문제를 해결하기 위한 전처리 모델과 염색체 검출을 위한 모델을 조합한 알고리즘이다.

2. 연구 방법

실험에 사용된 염색체 영상은 한국수력원자력(주) 방사선 보건원에서 획득한 영상으로, Axio Imager Z2 현미경과 Metafer 4 소프트웨어로 촬영한 염색체 영상을 사용한다. 전처리 작업 및 염색체 검출 라벨링은 전문가가 직접 색의 강도를 조정하고 염색체의 위치를 검출하였다.

그림 1은 제안하는 모델의 모식도를 보여준다. 제안하는 모델의 처리 과정은 먼저, FISH 염색된 안정형 염색체 이상 분석용 영상을 입력 영상으로 사용하고, 전처리 모델을 통해 색과 강도와 변경된 영상으로 변환한다. 다음으로 변환된 영상을 염색체 검출 모델의 입력 영상으로 사용하여 염색체를 개별로 검출하는 과정을 거친다.

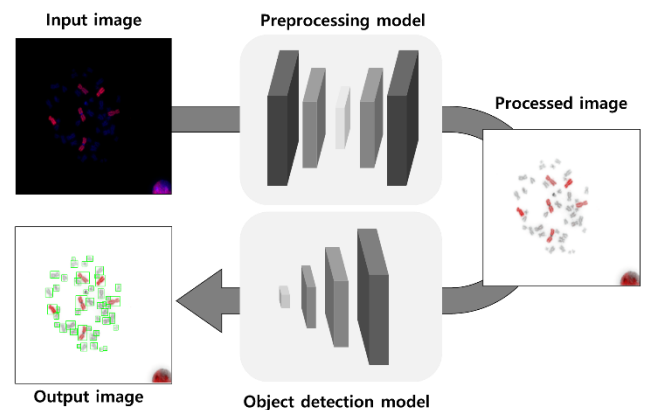


그림 1. 제안하는 모델 모식도

2.1 전처리 모델

전처리 과정은 FISH 염색 결과, 다른 색으로 염색된 1, 2, 4번 염색체의 색의 강도는 높고 그 외 염색체들의 색의 강도가 낮은 문제를 해결하기 위해, 기존에 전문가가 수동으로 색의 강도를 조정하는 과정을 딥러닝을 활용하여 자동화하였다. 전처리에 사용된 영상은 각각 554장으로 FISH 염색된 원본 영상과 전문가에 의해 직접 처리된 정답 영상으로 구성되어 있다. 또한, 사용된 영상의 양이 적기 때문에, 데이터 증강기법을 적용하였다. 데이터 증강기법은 무작위 상하 반전과 크기 조절을 적용하여, 매 반복마다

다른 영상으로 학습을 진행하였다.

전처리 작업에 사용된 모델은 U-Net[4] 모델을 기반으로 ResNet[5] 블록을 조합하여 사용하였다. 입력 영상과 출력 영상은 1024*1024*3 크기로 설정하였고, 형태는 그대로 유지하며 색의 강도의 변화만 기대하는 모델이기 때문에, mean square error가 상당히 낮게 계산되어 root mean square error를 손실함수로 사용하였고, 최적화 함수는 AdamW[6]를 사용하였다.

2.2 염색체 검출 모델

염색체 검출에 사용된 라벨 데이터는 총 1548장으로 전문가의 검수과정을 거쳐 각 영상의 염색체를 개별 분할하였다. 염색체 검출 모델 또한, 데이터의 양을 늘리기 위해 데이터 증강기법을 적용하였다. 사용된 데이터 증강기법은 무작위 위치 조정과 상하 반전, 일정 영역을 잘라 붙여 사용하는 모자이크 방법을 적용하였다.

제안하는 모델은 전처리 모델의 결과에 YOLOv5[6] 모델을 결합하여 구성하였고, 따라서 입력 영상의 크기는 전처리 모델의 출력크기로 설정하였다. 염색체 검출 모델에서 사용된 손실함수는 Binary cross entropy를, 최적화 함수는 AdamW를 사용하였다.

3. 연구 결과 및 고찰

전처리 모델에 사용된 554장의 영상은 8:2의 비율로 훈련과 검증 세트로 나누어 학습을 진행하였다. 그림 2은 전처리 모델의 결과를 비교한 그림이다. 그림 2a)는 FISH 염색된 안정형 염색체 이상 분석을 위한 영상이며 그림 2b)는 전문가에 의해 수동으로 색의 강도 조절이 된 전처리 모델 정답 영상이다. 그림 2c)는 전처리 모델의 출력 결과 영상으로 정답 영상과 상당히 유사한 것을 볼 수 있다. 그림을 보면, 모델 결과 영상의 색의 강도가 상대적으로 정답 영상보다 낮은 것을 확인할 수 있다. 추가로 다른 검증 영상을 비교해보면, 모델 결과 영상의 색의 강도가 정답 영상보다 상대적으로 높은 경우도 존재한다. 이는 정답 영상이 특정한 알고리즘으로 구성된 것이 아니라, 사람에 의해 수동으로 조정되어 과도하게 조정되거나 과소하게 조정된 경우가 발생하기 때문이다. 반면에, 모델 출력 결과는 평균 색의 강도가 유사하게 측정되어 과도하거나 과소하게 조정되지 않는 것을 확인할 수 있다.

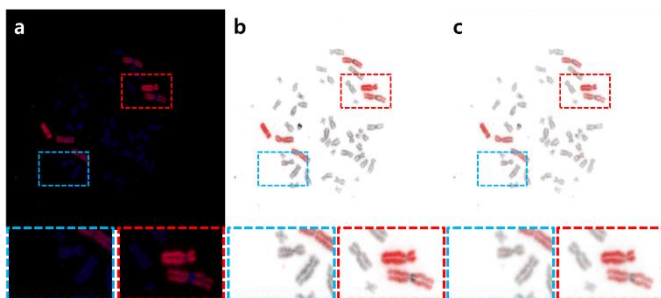


그림 3. 전처리 결과 비교

a) 원본 영상 b) 정답 영상 c) 모델 결과 영상

염색체 검출 모델에 사용된 영상도 8:2의 비율로 훈련과 검증 세트로 학습을 진행하였고, 그림 3에서 제안된 모델의 결과를 확인할 수 있다. 그림 3a)는 전처리 수행 전 원본 영상을 보여주고, 그림 3b)는 전처리 모델과 염색체 검출 모델이 모두 적용된 결과를 보여준다.

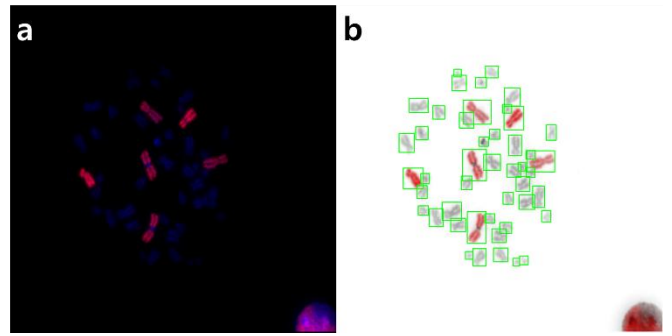


그림 2. 제안된 모델 결과

a) 원본 영상 b) 제안된 모델 결과

안정형 염색체 이상 분석은 불안정형 염색체 이상이 발견되지 않고, 안정형 염색체 이상만 발생하는 경우를 찾는 작업이기 때문에, 염색체의 개별 검출이 선행적으로 수행되어야 하는 작업이다. 하지만, 기존의 FISH 염색된 안정형 염색체 이상 분석용 영상은 1, 2, 4번의 염색체를 검출하기 위해 색의 강도가 과도하게 크게 나타나서 다른 염색체들이 상대적으로 색의 강도가 낮게 나타나고, 기존 영상을 활용한 모델의 결과도 좋지 않았다. 본 연구는 이러한 안정형 염색체 이상 분석을 위한 딥러닝 알고리즘을 처음으로 제시하였고, 그 결과로 제안된 모델이 개별 염색체 검출할 수 있다는 것을 확인하였다. 하지만 겹쳐진 염색체나 debris부분의 검출에서는 아직 정확도가 높지 않다는 한계점이 존재한다. 이러한 문제는 데이터를 추가하고 겹치는 염색체에 대한 개별 처리를 통해 해결할 수 있을 것으로 기대한다. 또한 추가적인 연구를 통해 안정형 염색체 이상 분석의 자동화도 기대할 수 있다.

4. Acknowledgements

이 연구는 한국수력원자력(주) 방사선 보건원의 지원을 받아 수행하였음(A211F03).

5.참고 문헌

[1] Patterson, D. Molecular genetic analysis of Down syndrome. *Human genetics*, Vol. 126, No.1, p.195-214, 2009

[2] Yan, J., Tucci, E., & Jaffe, N. Detection of t(9; 22) chromosome translocation using deep residual neural network. *Journal of Computer and Communications*, Vol.7, No.12, p.102-111, 2019

[3] Bai, H., Zhang, T., Lu, C., Chen, W., Xu, F., & Han, Z. B. Chromosome Extraction Based on U-Net and YOLOv3. *IEEE Access*, Vol.8, p.178563-178569, 2020

[4] Ronneberger, O., Fischer, P., & Brox, T. U-net: Convolutional networks for biomedical image segmentation. *In International Conference on Medical image computing and computer-assisted intervention p. 234-241, Springer, Cham, 2015*

[5] He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016

[6] Loshchilov, I., & Hutter, F. "Decoupled weight decay regularization." *arXiv preprint arXiv:1711.05101*, 2017