

암세포 영상분류를 위한 심층학습 모델 연구

박진형 · 최세운*

금오공과대학교

Deep Learning Model for Classification of Multiple Cancer Cell Lines

Jinhyung Park · Se-woon Choe*

Kumoh National Institute of Technology

E-mail : 20150555@kumoh.ac.kr / sewoon@kumoh.ac.kr

요 약

특정 질병 진단을 위한 병리 검사는 필수적이며, 최근 이러한 분야의 시간적, 인적 자원의 필요성을 줄이기 위해 인공 지능을 활용한 암세포의 자동분류가 가능한 시스템 구축에 관련된 연구가 활발하게 진행되고 있다. 하지만, 이전 연구에서는 제한적인 심층학습 알고리즘에 기인한 비교적 낮은 정확도로 데이터 처리에 한계가 존재하였다. 본 연구에서는 심층 학습의 일종인 Convolution Neral Network를 통해 4종류의 암세포를 4 Class Classification을 시행하는 방법을 제안한다. EfficientNet, ResNet, Inception을 사용하였으며 여러 하이퍼 파라미터 튜닝을 통해 얻은 모델을 앙상블 하여 최종적으로 97.26의 정확도를 얻을 수 있었다.

ABSTRACT

Additional pathological tests using imaging equipment are essential before diagnosing cancer cells. Recently, in order to reduce the need for time and human resources in these fields, research related to the establishment of a system capable of automatic classification of cancer cells using artificial intelligence is being actively conducted. However, in both previous studies, there were relatively limited deep learning algorithms and cell types, and limitations existed with low accuracy at the same time. In this study, a method of performing 4class Classification on four types of cancer cells through the Convolution Neral Network, a type of in-depth learning. EfficientNet, ResNet, and Inception were used, and finally Resnet was used to obtain an accuracy of 96.11 on average for k-fold.

키워드

Cancer Cell, Classification, CNN, Deep Learning

1. 서 론

암을 진단하기 위해서는 다양한 의료영상 장비를 사용하지만, 이는 체내 비정상적 조직의 존재유무, 위치, 크기 등과 같은 기본 정보의 제한적 제공이 가능하다. 따라서, 생리학적 정보를 획득하기 위해서는 다양한 추가 작업이 필요하다. 이를 위해 생검을 통해 조직을 획득한 후, 암 세포의 종류를 확인하기 위해 특정 파장에서 발현하는 형광물질 또는 특정 단백질에 바인딩이 되는 염색염료를 이용하여 조직을 특정화한다. 이후 광학 또는 형광 현미경을 이용하여 특정 암세포를 찾는 방식으로 생검조직을 이용한 확진이 이루어진다. 이러한 처리를 위해서 전문지식을 가진 종양학자, 임상병리

사 등의 도움과 생검조직의 처리를 위한 최소한의 소요 시간이 필요하며 또한 추가적인 비용이 발생하게 된다. 이러한 문제를 해결하기 위해 최근 인공지능을 활용한 암세포의 자동분류 시스템 구축에 관련된 연구가 활발하게 진행되고 있다 [1]. 하지만, 이전 연구들은 인공지능 학습을 위한 다양한 세포의 종류와 통계분석이 가능한 충분한 데이터 수집에 제한적이기 때문에 정확도의 한계가 존재한다 [2-4]. 따라서, 본 연구에서는 심층학습의 일종인 합성곱 신경망을 이용한 4가지 암세포의 현미경 이미지를 식별하는 방법을 제안하고자 한다.

* corresponding author

II. 실험 방법

2.1 이미지 획득

본 연구에서는 다양한 암 세포들을 배양하고, 현미경을 통해 세포 광학 이미지를 데이터를 수집하였다. 실험을 위해 배양한 세포는 Hela (cervical cancer cells), MCF-7 (breast cancer cells), Huh7 (liver cancer cells), NCI-H1299 (lung cancer cells)이며, 각각의 세포는 5%의 이산화탄소를 포함한 CO2 Incubator를 사용하여 37°C의 온도에서 배양되었다. 세포배양을 위해 High concentration glucose를 포함한 Dulbecco's Modified Eagle Medium (DMEM), 10%의 Fetal Bovine Serum (FBS, pH 7.4)과 1%의 Penicilin Streptomycin으로 만들어진 배지 용액을 사하여 배양하였다. 세포들은 3-5일간 배양하면서 광학현미경 (Inverted optical microscope, IX73, Olympus, Japan)을 사용하여 이미지 획득을 진행하였다.

2.2 이미지 전처리

세포의 이미지를 인공지능에 사용하기 위해서, 현미경 세포 이미지의 전체 이미지를 학습하는 것이 아니라 세포가 있는 부분만 이미지를 segmentation 하여 학습을 진행하였다. 이 과정을 위해 Python(3.6.9)의 package 중 openCV (4.5.1.48)와 scikit-image(0.17.2)를 이용하여 이미지 전처리를 하였다. 간략하게 과정을 설명하자면, openCV를 이용하여 threshold를 조절하고 scikit-image를 통해 배경을 제거하는 등의 전처리를 통해 세포의 윤곽선을 찾고 각각의 세포를 segmenation하여 저장하는 과정을 진행한다. 선별된 세포 이미지 중 부유물, 이물질, 죽은 세포, 미숙 세포 등과 같이 가로세로 픽셀의 합이 100 이하인 이미지는 노이즈로 가정하고 정확도 증가를 위해 제외하였으며, 학습의 용이성을 위해 모든 이미지는 200 × 200 pixel 사이즈로 조절하였다. 이미지 처리 후 가장 적은 암세포 이미지 데이터를 기준으로 4개의 세포그룹 이미지를 통일하였고, 총 27,000장을 train, validation, test 그룹으로 나누었다. 데이터는 3:1:1의 비율로 5-fold를 사용하여 나누었으며 학습에 사용된 데이터는 16,200장으로 각 세포 당 4,050장의 train data set으로 학습이 진행되었다.

III. 실험 결과

모델 학습은 Tensorflow(2.4.1)안의 Keras를 통해 진행되었다. 컴퓨터 환경은 Ubuntu 18.04.5 LTS, CPU Intel i9-10940X, GPU GeForce RTX 3090×2이며 Cuda는 11.2로 학습이 진행 되었다.

Keras API를 통해 제공되는 pre-trained model을 사용하여 학습이 진행 하였으며 이전 실험에서 Optimizer와 Model을 결정하였다. 모델로는

Inception, Resnet, EfficientNet 총 세가지 모델을 사용하였고, optimizer는 Adagrad, SGD를 사용하여 비교 실험을 진행하였다. 모델 학습을 위한 새로운 classifier를 재정의하였고, 학습에 여러 가지 기법들을 비교하여 최적의 모델을 찾아가는 실험을 진행하였다. 그리고 다양한 조건을 활용한 비교 실험을 진행하기 위해서 K-fold를 사용하여 여러 모델을 같은 조건으로 학습하고 비교함으로써 정확도를 증가시켰다. 예를들어, learning rate, augmentation과 layer freeze 등의 조건을 ON/OFF 방식으로 인가하여 학습을 진행하였다. Learning rate는 exponential learning rate decay를 사용하여 학습을 시켰다. Augmentation을 shift와 flip, rotation이 random하게 인가될 수 있도록 만들어 주었다. 일반적으로 domain이 같은 경우, feature extraction을 담당하는 전반부를 freeze하고 후반부나 classifier만을 학습하는 방식을 취하지만, domain이 다를 경우 layer 전체를 학습하는 방식을 취한다. ImageNet의 데이터와 세포 이미지는 자연 이미지와 광학 이미지로 매우 다른 domain을 가지고 있으므로 전체를 train 하는 것이 일반적이나, freeze의 경우 속도적인 측면과 overfitting을 비교적 극복하는 경향이 있기에 초반 25%를 학습하여 비교하는 실험을 진행하였다. 결론적으로 learning rate decay, augmentation, layer freeze 등과 같은 세부 조건을 적용했을 때 비교적 높은 정확도를 얻을 수 있었다. 표 1에서 이러한 조건에서의 여러 모델의 결과를 확인 할 수 있다. Resnet5와 Adagrad를 사용 하였을 때 96.26%의 k-fold평균 Accuracy를 얻을 수 있었다.

표 1. Model 비교 결과

Model	Optimizer	Accuracy
EfficientNetB2	SGD	95.70%
EfficientNetB2	Adagrad	95.95%
InceptionV3	SGD	95.56%
InceptionV3	Adagrad	95.20%
Resnet50	SGD	96.11%
Resnet50	Adagrad	96.26%

IV. 결론

본 연구에서는 심층학습의 일종인 합성곱 신경망을 이용한 4가지 암세포의 현미경 이미지를 식별하는 방법을 제안하고, 다양한 조건을 사용하여 학습한 후 테스트를 진행하여, 그 효과를 분석하였다. 그 결과 제안된 방식을 통한 결과는 비교적 높은 정확도를 보였으며, 향후 추가 연구를 통해 높은 정확도와 함께 빠른 처리속도를 지향하는 최적화 연구를 진행할 예정이며, GAN과 같은 생성 모델을 사용할 예정이다.

Acknowledgement

이 논문은 2020년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2019R1F1A1062397). 본 논문은 4단계 BK21 사업(금오공과대학교 IT융복합학공학과)에 의하여 지원되었음.

References

- [1] R. Wang, D. Wang, D. Kang, X Guo, C. Guo, M. Dongye, Y. Zhu, C. Chen, X. Zhang, E. Long, X. Wu, Z. Liu, D. Lin, J. Wang, K. Huang and H. Lin, “An artificial intelligent platform for live cell identification and the detection of cross-contamination,” *Annals of Translational Medicine*, Vol. 8, No. 11, 697, June 2020.
- [2] S.H. Kim, J.H. Lee, E.Y. Choi, S.T. Jeon, M.Y. Choi, S.H. Jo and S. Choe, “Automatic cell image classification with convolutional neural networks,” *The Transactions of the Korean Institute of Electrical Engineers*. Vol. 70, No. 1, 139-144, Jan. 2021.
- [3] G. Ayana, K. Dese and S. Choe, “Transfer Learning in Breast Cancer Diagnoses via Ultrasound Imaging,” *Cancers*, Vol. 13, No. 4, 738, Feb. 2021.
- [4] K. Cho and S. Choe, “LED를 활용한 저가의 암세포 증식제어 모듈 개발 및 효과,” *한국정보통신학회논문지*, Vol. 22, No. 9, pp. 1237 - 1242, Sep. 2018.
- [5] K. Cho, J.-h. Seo, G. Heo and S. Choe, “An Alternative Approach to Detecting Cancer Cells by Multi-Directional Fluorescence Detection System Using Cost-Effective LED and Photodiode,” *Sensors*, Vol. 19, No. 10, pp. 2301, May 2019