

심층 학습을 통한 암세포 광학영상 식별기법

박진형 · 최세운*

금오공과대학교

Identification of Multiple Cancer Cell Lines from Microscopic Images via Deep Learning

Jinhyung Park · Se-woon Choe*

Kumoh National Institute of Technology

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

요 약

임상에서 암 관련 질병의 확진을 위해 영상장비를 이용한 기초 진단 이후 추가적인 방법으로 생체검사 등을 이용한 병리적 검사가 필수적이다. 이러한 생체검사를 진행하기 위해서는 전문지식을 가진 종양학자, 임상병리사 등의 도움과 최소한의 소요시간은 확진을 위해 반드시 필요하다. 최근 들어, 인공지능을 활용한 암세포의 자동분류가 가능한 시스템 구축에 관련된 연구가 활발하게 진행되고 있다. 하지만, 이전 연구들은 한정된 알고리즘을 기반으로 하여 세포의 종류와 정확도에 한계를 보인다. 본 연구에서 심층 학습의 일종인 합성곱 신경망을 통해 총 4가지의 암세포를 식별하는 방법을 제안한다. 세포 배양을 통해 얻은 광학영상을 OpenCV를 사용하여 세포의 위치 식별 및 이미지 분할과 같은 전처리 수행 후, EfficientNet을 통해 학습하였다. 모델은 EfficientNet을 기준으로 다양한 hyper parameter를 사용하고, InceptionV3을 학습하여 성능을 비교분석 하였다. 그 결과 96.8%의 높은 정확도로 세포를 분류하는 결과를 보였으며, 이러한 분석방법은 암의 확진에 도움이 될 것으로 기대한다.

ABSTRACT

For the diagnosis of cancer-related diseases in clinical practice, pathological examination using biopsy is essential after basic diagnosis using imaging equipment. In order to proceed with such a biopsy, the assistance of an oncologist, clinical pathologist, etc. with specialized knowledge and the minimum required time are essential for confirmation. In recent years, research related to the establishment of a system capable of automatic classification of cancer cells using artificial intelligence is being actively conducted. However, previous studies show limitations in the type and accuracy of cells based on a limited algorithm. In this study, we propose a method to identify a total of 4 cancer cells through a convolutional neural network, a kind of deep learning. The optical images obtained through cell culture were learned through EfficientNet after performing pre-processing such as identification of the location of cells and image segmentation using OpenCV. The model used various hyper parameters based on EfficientNet, and trained InceptionV3 to compare and analyze the performance. As a result, cells were classified with a high accuracy of 96.8%, and this analysis method is expected to be helpful in confirming cancer.

키워드

Cancer Cell, Classification, CNN, Deep Learning

1. 서 론

일반적으로, 암을 진단하기 위해서는 X-ray, 초

음파 촬영 또는 CT 등을 실시하여 진단한다. 하지만 이는 체내 비정상적 물체의 존재, 위치, 크기 등을 알기만 할 뿐, 이의 원인이 암인지, 어떤 암인지 확인할 수 없다. 이를 위해 해당 종양의 조각을 채취 후 세포를 배양하고 현미경으로 검사하여

* corresponding author

암세포를 찾는 생검 방식으로 암진단이 이루어진다. 이러한 생체검사를 하기 위해서는 전문지식을 가진 종양학자, 임상병리사 등의 도움과 최소한의 소요 시간은 확진을 위해 필요하다. 이러한 문제를 해결하기 위해 최근 인공지능을 활용한 암세포의 자동분류 시스템 구축에 관련된 연구가 활발하게 진행되고 있다 [1]. 하지만, 이전 연구들은 세포의 종류가 다양하지 못하거나 [2] 다양하더라도 전체를 다 같이 비교하는 것이 아닌 세포를 이진 분류를 하는 형태로 대부분 이루어 졌으며 정확도에 한계를 보였다 [3]. 우리는 본 연구를 통해 심층 학습의 일종인 합성곱 신경망을 이용하여 총 4가지의 암세포 현미경 이미지를 식별하는 방법을 제안한다.

II. 데이터 수집

본 논문에서는 자체적으로 다양한 세포를 배양하며 동시에 광학 이미지를 획득하여 데이터를 수집하였다. 실험에 사용된 세포는 총 네 가지이며 Hela(human, female, 31 years old, Black, cervical cancer cells), MCF-7(human, female, 69 years old, Caucasian, breast cancer cells), Huh7(human, liver cancer cells), NCI-H1299(human, lung cancer cells)의 세포로 구성된다. 각각의 세포는 전부 37°C의 온도와 5%의 이산화탄소를 포함한 CO₂ Incubator에서 배양하였으며, 고농도의 글루코스가 포함된 Dulbecco's Modified Eagle Medium (DMEM), 10%의 Fetal Bovine Serum (FBS, pH 7.4)과 1%의 Penicilin Streptomycin으로 구성된 배지 용액으로 3-4일 배양 후 적절히 자란 세포들을 역상 광학현미경 (Inverted optical microscope, IX73, Olympus, Japan)을 이용하여 촬영하였다.

III. 실험 결과

3.1 OpenCV 활용 이미지 전처리

세포 배양을 통해 얻은 이미지를 바로 학습을 하기에는 무리가 있다. 현미경 세포 이미지의 전체 이미지를 학습하는 것이 아니라 세포가 있는 부분만 이미지를 분리하여 학습을 진행하였다. 이 과정에서 조금 더 효과적으로 세포를 Segmentation을 하기 위해 Python(3.6.9)의 package 중 openCV(4.5.1.48)와 scikit-image(0.17.2)를 이용하여 이미지 전처리를 하였다. openCV를 이용하여 threshold를 조절하고 scikit-image를 통해 배경을 제거하는 등의 전처리를 통해 세포의 윤곽선을 찾고 각각의 세포를 segmenation하여 저장하는 과정을 진행한다. 이때, 세포 이미지 내의 작은 부유물, 이물질, 죽은 세포, 미숙 세포 등을 제외하

기 위해 가로세로 픽셀의 합이 100 이하인 이미지는 제외하였으며 학습의 용이를 위해 모든 이미지는 200 × 200의 이미지로 크기를 조절하였다. 이미지 처리 후의 데이터의 이미지 숫자 중 가장 적은 데이터의 숫자에 맞추어 데이터의 숫자를 균일하게 만들었고 총 27,000장을 train, validation, test 그룹으로 나누었다, 데이터는 7:2:1의 비율로 나누었으며 학습에 사용된 데이터는 18,900장으로 각 세포 당 4,725장의 train data set으로 학습이 진행되었다.

3.2 모델 학습 및 비교

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.

Keras에서는 ImageNet으로 pre-train을 통한 다양한 모델을 제공한다. 본 실험에서는 이런 모델을 로드하여 진행되었으며, 가장 먼저 세포 단위의 딥러닝에서 효과적인 optimizer를 찾아내는 실험을 진행하였다. 기본이 되는 모델 EfficientB3로 선택하였고, input은 200 × 200이고 Global Average Pooling을 진행 후에 softmax를 통해 4진 분류 결과를 도출하는 형식으로 진행하였다. Optimizer는 총 8가지를 비교하였으며 결과는 표 1과 같다. 해당 표의 Loss와 Accuracy는 테스트 데이터를 통해 얻은 test loss와 test accuracy이다. 결과를 보면 다른 optimizer와는 달리 Ftrl의 경우 학습효과가 상대적으로 낮음을 볼 수 있다. Optimizer에 대한 비교는 세포 단위의 딥러닝에서 효과가 있는 optimizer를 찾는 것이 목표였기 때문에, 결과적인 정확도 보다는 모델의 학습곡선의 경향을 보며 optimizer를 선정하였다. 결과적으로 Optimizer는 SGD, Adagrad, Adadelat를 통해 실험을 계속해서 진행 하였다.

Keras에서 쉽게 사용할 수 있는 모델 중, 정확도가 높을 것으로 예상되는 모델을 논문 리뷰를 통해 선별하고, 정확도를 비교하는 실험을 진행하였다. ImageNet으로 pre-train된 모델 중 3가지 모델을 선택하여 진행하였으며, 복수의 모델 버전을 선택하여 추가로 진행하였다. 학습에 사용된 모델은 VGG16, VGG19, InceptionV3, EfficientNetB2, EfficientNetB3 총 5가지의 모델을 사용하여 비교 하였다. 모델별 정확도 실험은 optimizer 비교 실험과 동일한 조건에서 진행되었으며, 가장 흔히 사용되는 Adam으로 비교 실험을 진행하였다. 표2에서 확인할 수 있듯이, Loss와 Accuracy는 test의 결과이며 VGG 모델보다 InceptionV3, EfficientNetB2, EfficientNetB3 의 세 가지 모델이 보다 높은 정확도를 보이는 것을 확인할 수 있었다.

위와 같은 실험을 통해, 4가지 서로 다른 종류의 암세포를 보다 정확하게 분류할 수 있는 Model로 InceptionV3와 EfficientNetB2를 선정하였는데

EfficientNetB3와 비교해 EfficientNetB2가 속도 면에서 더 우수하였기에 B2로 선정하여 실험을 진행하였다. Optimizer는 위에서 설명한 SGD, Adagrad, Adadelta를 통해 실험을 진행하였다. 추가적으로 180도 이내에서 무작위로 회전하는 Augmentation 프로세스를 진행하여, train 데이터만 이미지의 숫자를 증가시켜 실험을 진행하였으며 그로 인해 전체적으로 정확도가 상승한 것을 확인할 수 있었다. 이러한 실험을 통해 선택된 조건을 포함하여 Model 2개 optimizer 3개 총 6개의 조합을 학습시켰으며, 학습 결과 EfficientNetB2와 InceptionV3 모두 SGD와 함께 사용하였을 때 가장 좋은 결과를 얻을 수 있었으며, Adagrad, Adadelta 순으로 두 모델 모두 Accuracy의 순위는 같았다. 결과적으로 EfficientNetB2와 SGD의 조합이 가장 높은 정확도인 96.87%의 결과를 얻을 수 있었다.

표 1. Optimizer 비교 결과

Optimizer	Loss	Accuracy
Adadelta	0.44	85.55%
Adagrad	0.25	92.19%
Adamax	0.44	93.36%
Ftrl	1.39	28.52%
Nadam	0.5	89.84%
RMSprop	0.59	91.02%
SGD	0.31	90.62%
Adam	0.36	91.02%

표 2. Model 비교 결과

Model	Loss	Accuracy
VGG16	1.44	28.26%
VGG19	1.39	23.44%
EfficientNetB2	0.47	90.62%
EfficientNetB3	0.36	91.02%
InceptionV3	0.43	92.58%

표 3. 최종 비교 결과

Model	Opimizer	Loss	Accuracy
EfficientNetB2	SGD	0.08	96.87%
	Adagrad	0.14	94.92%
	Adadelta	0.20	91.40%
InceptionV3	SGD	0.18	95.70%
	Adagrad	0.23	94.53%
	Adadelta	0.18	92.96%

IV. 결 론

본 연구를 통하여 Keras에서 제공하는 모델을 적절히 최적화하면 다양한 암세포의 분류를 비교적 높은 정확도로 가능하다는 것을 확인하였다. 이번 실험은 모델을 비교하고 정확도를 높이는 optimizer의 선택을 위한 실험을 대부분 진행하였고, 학습된 모델을 통해 결과를 내는 prediction 소요시간은 5~7초 가량이 소요됨을 확인할 수 있었다. 향후 추가연구를 통해 세포 이미지를 증가시켜 신뢰도의 향상 등 실제로 암 진단에 도움이 될 수 있도록 관련 연구를 진행할 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.-w. 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.-w. Choe, "Transfer Learning in Breast Cancer Diagnoses via Ultrasound Imaging," *Cancers*, Vol. 13, No. 4, 738, Feb. 2021.