Final Project : Simple Classification of Malignant and Benign Skin Cancers Using CNN

2015312904, Dept. of Electronic and Electrical Engineering, Joon Woo Kwon

Abstract

피부암은 피부에 발생하는 악성 종양을 총칭하는 말로, 2018년엔 전세계적으로 130만건 이상의 새로운 피부암 진단이 이루어질 정도로 세계에서 가장 흔한 형태의 암 종류 중 하나입니다. 피부암은 보통 피부과 전문의의 진찰을 통해 시각적으로 발견됩니다. 이런 진단 특성을 토대로 본 보고서에선 컨벌루션 뉴럴 네트워크 (CNN)를 사용하여 이미지를 통해 양성 및 악성 피부 종양을 구분하는 단순 네트워크 모델을 구축하고 이 모델이 실제로 얼마만큼의 정확도를 가지는 지 분석합니다.

Keyword: Convolutional Neural Networks, Skin Cancer, Image Classification, Machine Learning, Deep Learning

1. Introduction

1.1. Background

1.1.1. 피부암 (Skin Cancer)

피부암은 피부조직을 구성하는 세포의 조절되지 않는 비정상적인 과증식, 즉 악성화한 것을 총칭합니다. 피부암은 전체 악성 종양의 40%를 차지하는 가장 흔히 발병하는 암이며 피부암의 80%는 얼굴, 머리, 손목 등 태양광에 노출된 부위에서 일어납니다.

동양인에서의 발생 빈도는 서양인에 비해서 적지만 자외선 노출 정도가 커지고, 피부에 각종 유해 물질의 노출 기회가 증가됨에 따라 우리나라에서도 피부암의 발생빈도가 증가하는 추세입니다.

통계에 따르면 최근 5년 사이에도 국내 피부암 환자가 44%나 급증한 것을 확인할 수 있습니다. 대한피부과학회가 최근 발표한 피부암 환자 증가 추이 분석에 따르면 2009년 1만 989명이던 피부암 환자가 1만5천826명으로 한 해 평균 9.6%, 5년간 총 44.1% 증가한 것으로 조사됐습니다.

1.1.2. 피부암 조기 진단의 필요성

피부는 통각, 촉각, 압각, 온도 등을 느끼게 합니다. 그리고 혈관 확장, 수축을 통해 체온을 조절하고 세균이나 화학물질 등의 이물질이나 열, 자외선으로부터 신체를 보호해 줍니다. 그런데 피부에 암이 발생하면 우리 몸에 반드시 필요한 이러한 기능을 해주지 못합니다.

피부암의 종류는 매우 다양하며, 피부의 이상은 내부 장기의 경우와는 달리 눈으로 보아 비교적 쉽게 알 수 있으므로, 조기에 발견할 가능성이 높다고 할 수 있으며, 적절한 시기에 치료하면 대부분 완치 가능한 질환입니다. 그러나 악성 피부암은 예후가 좋지 않으므로 초기에 정확히 진단하는 것이 가장 중요합니다.

따라서 피부과 전문의에 의한 피부암의 조기 진단은 피부암의 예방을 위해 필수적이라고 할 수 있습니다.

1.1.3. The Objective

피부암은 앞서 언급한 바와 같이 피부과 전문의의 Visual Diagnosis로 쉽게 detecting이 가능합니다. 하지만 시간적, 공간적 이유로 바쁜 현대인들이 피부과 전문의에게 피부암으로 의심이 가는 부위를 바로 진단 받기란 쉽지 않습니다.

이러한 진단 특성 및 배경 위에서 본 보고서는 딥러닝 기법 중 하나인 컨벌루션 뉴럴 네트워크 (CNN)을 사용하여 이미지를 통해 양성 및 악성 피부 종양을 구분하는 단순 네트워크 모델을 구축하고 이 모델들이 실제로 얼마만큼의 정확도를 가지는 지 분석합니다. 구축한 모델이 일정 수준 이상의 정확도를 가진다면 피부암 초기 진단에 본 모델이 더 저렴하고 더 정확한 진단을 하는 것에 큰 역할을 할 수 있을 것입니다.

2. The Dataset and Algorithms

2.1. Description of Dataset

위 문제를 해결하기 위해서 ISIC (International Skin Image Collaboration)-Archive의 skin cancer dataset을 가져왔습니다.
(All the rights of the Data are bound to the ISIC-Archive rights (https://www.isic-archive.com/#!/topWithHeader/wideContentTop/main).)

이 dataset은 1800장의 양성 종양 사진과 1497장의 악성 종양으로 분류된 사진들로 구성되어 있습니다. 또한 노트북 컴퓨터의 한계로 모든 사진들을 원본으로 돌릴 수가 없었기에 사진은 모두 저해상도 (224x224x3) RGB로 크기가 조정되었습니다. Dataset에서 random하게 악성 및 양성 종양 사진을 출력하면 아래와 같습니다.

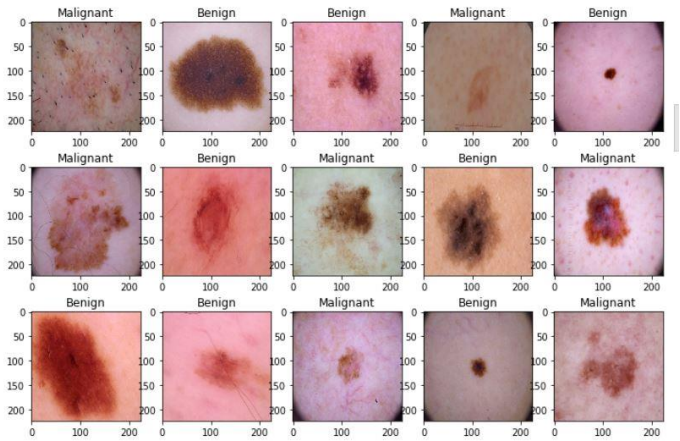

Fig.1. The picture of Dataset (Benign & Malignant)

2.2. Convolutional Neural Network (CNN)

2.2.1. The algorithm

하지만 단순히 피부암 이미지를 사용하여 해당 피부 병변이 피부암인지 아닌지를 자동으로 분류하는 것은 피부 병변 모양이 다 똑같지 않고 세밀하게 다르기 때문에(피부 병변의 가변성) 굉장히 어려운 작업입니다.

따라서 본 보고서에선 컨벌루션 뉴럴 네트워크(CNN)을 사용하여 이 문제를 해결하고자 했습니다.

2.2.2. Why CNN?

컨벌루션 뉴럴 네트워크(CNN)는 모델이 직접 이미지, 비디오, 텍스트 또는 사운드를 분류하는 머신러닝의 한 유형인 딥러닝에 가장 많이 사용되는 알고리즘입니다.

CNN은 특히나 이미지에서 객체, 얼굴, 장면을 인식하기 위해 패턴을 찾는데 특히 유용합니다. CNN은 데이터에서 직접 학습하며, 패턴을 사용하여 이미지를 분류합니다. CNN은 특징을 직접 학습하기 때문에 특징을 수동으로 추출해야 할 필요가 없으며 현재 다양한 딥러닝 Algorithm 중에 가장 높은 수준의 인식 결과를 보입니다. 또한 기존 네트워크를 바탕으로 한 새로운 인식 작업을 위해 CNN을 재학습하여 사용하는 것이 가능합니다.

위와 같은 장점으로 인해 자율 주행 자동차, 얼굴 인식 어플리케이션과 같이 객체 인식과 컴퓨터 비전이 필요한 분야에서 CNN은 널리 사용됩니다. CNN의 Rough한 작동 원리는 아래의 figure에 잘 묘사 되어 있습니다.

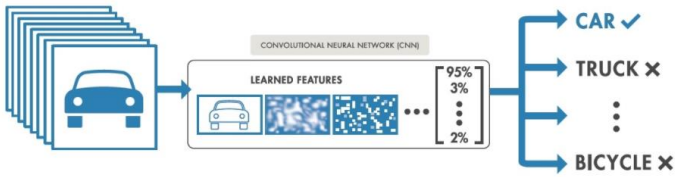

Fig.2. CNN에 이미지가 전달되면 자동으로 특징을 학습하고 객체를 분류합니다.

2.2.3. CNN의 작동원리

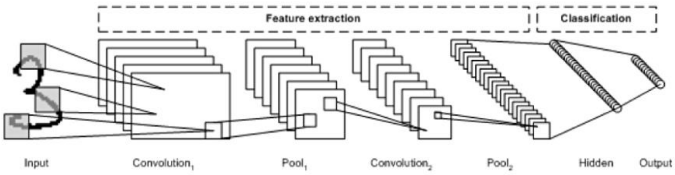

Fig.3. The structure of CNN

CNN은 위 이미지와 같이 이미지의 특징을 추출하는 부분과 클래스를 분류하는 부분으로 나눌 수 있습니다. 특징 추출 영역은 Convolution Layer와 Pooling Layer를 여러 겹 쌓는 형태로 구성됩니다. 또한 CNN 마지막 부분에는 이미지 분류를 위한 Fully Connected 레이어가 추가됩니다.

CNN은 이미지 특징 추출을 위해 Fig.4.와 같이 입력데이터를 필터가 순회하며 합성 곱을 계산하고, 그 계산 결과를 이용하여 Feature map을 만듭니다.

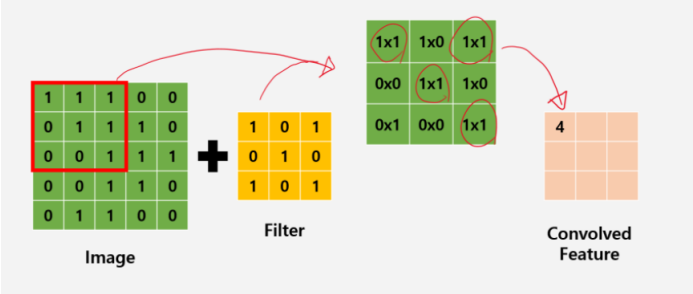

Fig.4. Convolutional Neural Network

또한 Pooling 레이어를 통해 이미지의 크기를 적당하게 줄여 연산량을 줄이고, Feature map의 특정 feature를 강조합니다. CNN에서는 주로 Max Pooling을 사용하는데 이는 뉴런이 가장 큰 신호에 반응하는 것과 유사하다고 합니다.

이러한 Convolution과 Pooling을 반복적으로 사용하는 Feature extraction을 통해 각 이미지의 불변하는 특징을 찾고 이 특징을 입력데이터로 Fully-Connected Layer에 보내 Classification을 수행합니다. 따라서 CNN을 이용하면 피부 병변 사진을 보고 시각적으로 양성인지, 악성 종양인지를 더 정확하게 분류할 수 있을 것입니다.

3. Experiments and Remarks

3.1. CNN models with keras & TensorFlow (LeNet)

3.1.1. The parameters

모든 시뮬레이션은 Google Colab과 Intel® Core ™ i5-5200U CPU @ 2.20Ghz and 8GB of memory 노트북 컴퓨터로 실행되었습니다.

CNN 모델에 사용된 CNN 구조와 각종 변수는 아래의 Fig.5.와 Table 1. 과 같습니다. Test set과 Train set은 8:2의 비율로 구분되었습니다.

또한 케라스 콜백의 ReduceLROnPlateau 함수를 이용하여 validation accuracy가 5번 동안 줄지 않으면, 학습률을 1/2로 줄이고 줄여도 minimum learning rate가 1e-7보다 작아지지 않게 설계하였습니다.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 224, 224, 64)      1792
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 112, 112, 64)      0
_________________________________________________________________
dropout (Dropout)            (None, 112, 112, 64)      0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 112, 112, 64)      36928
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 56, 56, 64)        0
_________________________________________________________________
dropout_1 (Dropout)          (None, 56, 56, 64)        0
_________________________________________________________________
flatten (Flatten)            (None, 200704)            0
_________________________________________________________________
dense (Dense)                (None, 128)               25690240
_________________________________________________________________
dense_1 (Dense)              (None, 2)                 258
=================================================================
Total params: 25,729,218
Trainable params: 25,729,218
Non-trainable params: 0
_________________________________________________________________
```

Fig.5. CNN used in experiments

Table 1. The parameters

Kernel_size	(3,3)
Dropout	0.25
Learning rate	1e-3 & ReduceLROnPlateau
Number of classes	2
Activation Function	Relu
Optimizer	Adam
Loss function	Binary_crossentropy
Metrics	Accuracy
Pooling	Max-pooling

3.1.2. The Result

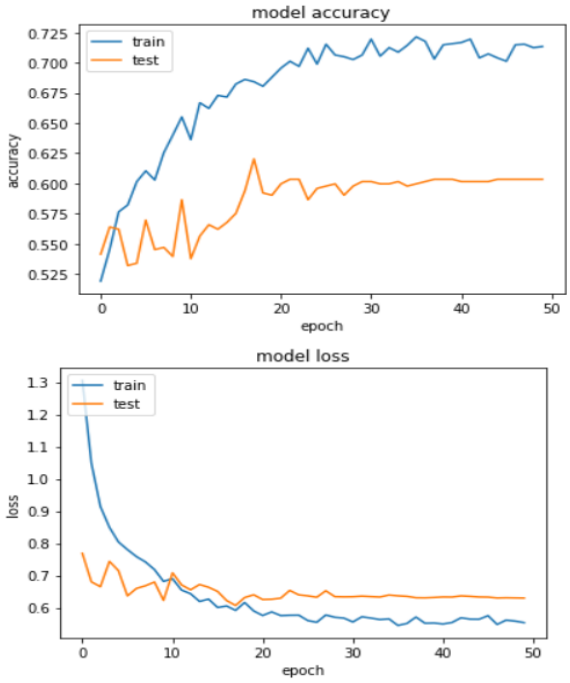

Fig.6. The result of training set for LeNet

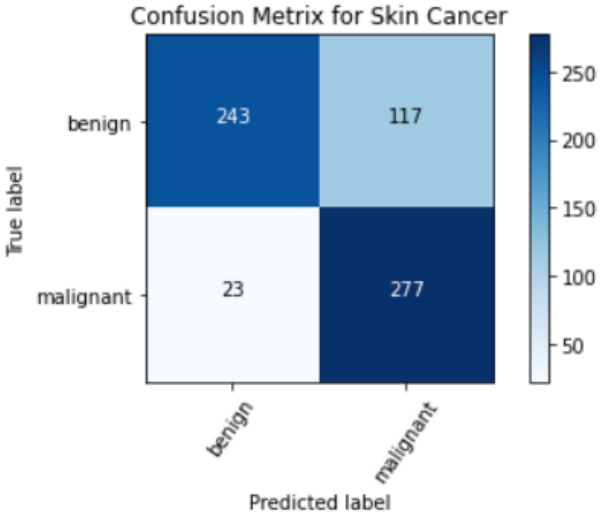

Fig.7. The test set Confusion Matrix for LeNet

3.2. ResNet50

3.2.1. The parameters

이번에는 ResNet50을 이용하여 실험을 진행하였으며 실험에 쓰인 ResNet50의 parameters 및 구조는 아래와 같습니다.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 64 \\ 3\times3, 64 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 64 \\ 3\times3, 64 \\ 1\times1, 256 \end{bmatrix}\times3$
conv3_x	28×28	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 128 \\ 3\times3, 128 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times4$	$\begin{bmatrix} 1\times1, 128 \\ 3\times3, 128 \\ 1\times1, 512 \end{bmatrix}\times8$
conv4_x	14×14	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 256 \\ 3\times3, 256 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times6$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times23$	$\begin{bmatrix} 1\times1, 256 \\ 3\times3, 256 \\ 1\times1, 1024 \end{bmatrix}\times36$
conv5_x	7×7	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times2$	$\begin{bmatrix} 3\times3, 512 \\ 3\times3, 512 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$	$\begin{bmatrix} 1\times1, 512 \\ 3\times3, 512 \\ 1\times1, 2048 \end{bmatrix}\times3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

Fig.8. The structure of ResNet50

또한 레이어가 50개 이상인 ResNet은 아래와 같은 Bottleneck skip connection 구조를 사용합니다.

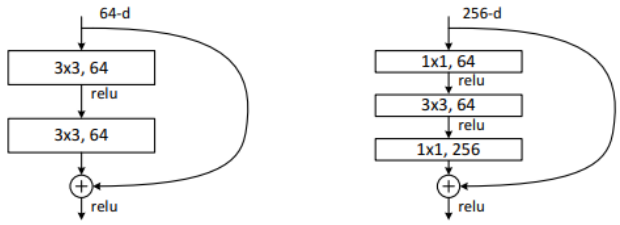

Fig.9. The bottleneck skip connection of ResNet

Table 2. The parameters used in Resnet50

Kernel_size	First Layer: (7x7) The Rest: (3,3)
Model	Resnet50
Learning rate	1e-5 & ReduceLROnPlateau
Number of classes	2
Activation Function	Relu
Optimizer	Adam
Loss function	Binary_crossentropy
Metrics	Accuracy
Pooling	Max-pooling avg-pooling

3.2.2. The result

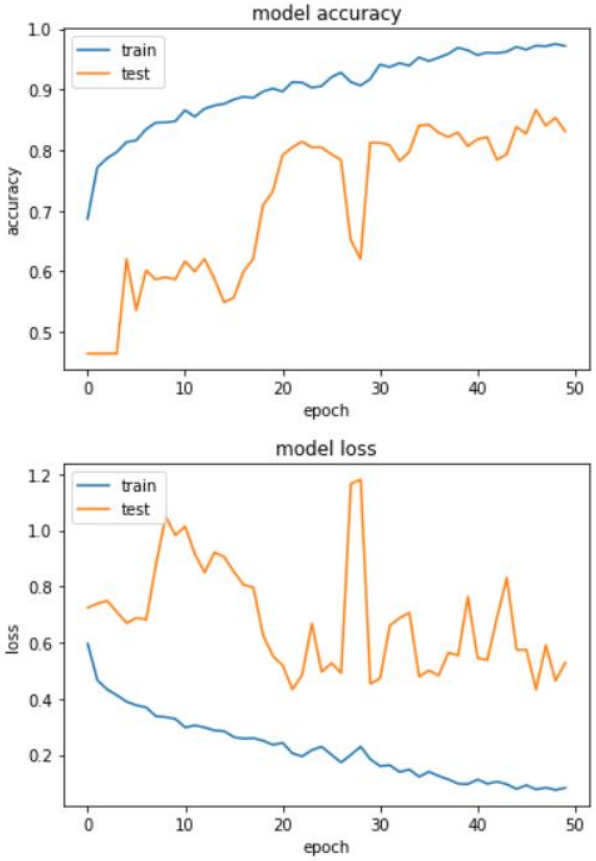

Fig.10. The result of training set for Resnet50

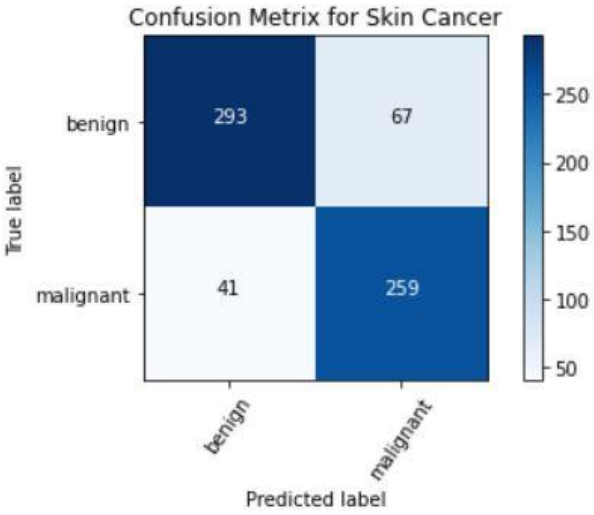

Fig.11. The test set Confusion Matrix for ResNet50

3.3. Remarks

Table 3. The Test set confusion matrix for each model

Model	Confusion Matrix
LeNet	[243 117] [23 277]
ResNet50	[293 67] [41 259]

Table 4. Test set accuracy comparison (Score)

Model	Accuracy
LeNet	0.79
ResNet50	0.83

Table 3와 Table 4를 통해 CNN을 이용한 종양의 종류 예측이 성공적으로 이루어졌음을 확인할 수 있었으며 두 모델 다 Accuracy가 80% 근방이라는 점에서 좋은 성능을 보였다는 것을 확인할 수 있습니다.

결과를 보면 ResNet50이 LeNet보다 Accuracy가 근소하게 높은 것을 확인할 수 있었습니다. Resnet50 모델의 Accuracy가 0.83 이라는 것은 이 모델이 테스트된 전체 종양의 83%를 양성 종양인지 악성 종양인지 정확하게 예측했다는 뜻입니다. 이는 처음 위 실험에서 제시한 문제의식인 종양 종류 파악에 큰 도움이 될 수 있는 수치이며 실제 산업에서 지금 바로 쓰여도 큰 활약을 할 수 있을 것이라고 생각합니다.

허나 LeNet 같은 경우 처음에 Training 할 때 trainingset의 accuracy와 validation set의 accuracy가 Resnet50 보다 크게 차이가 나는 것을 확인할 수 있습니다. 이는 아마도 dataset을 random하게 8:2의 비율로 train set과 test set으로 나누었고 그리고 train set 안에서도 8:2로 validation set을 나눴기 때문에 거기서 비롯되는 문제라고 생각하지만 후속 연구가 필요할 것 같습니다.

또한 learning rate의 한계를 학습의 속도가 너무 느려 1e-7로 고정해두었는데 이를 조정하거나 또는 Loss Function을 조정해보면서 다양한 실험을 cast해도 괜찮을 것 같습니다.

4. Conclusions and Future Works

위 보고서에서 우리는 피부암과 피부암 조기 진단의 필요성 그리고 위 상황에서 왜 CNN이 쓰여야 하는지에 대해 간단히 리뷰한 뒤, 양성 및 악성 피부암을 잘 구분하는 모델을 1800장의 양성 종양 사진과 1497장의 악성 종양 사진 데이터에 LeNet 및 ResNet과 같은 다양한 CNN Model을 적용하여 설계하였습니다.

추가로 위 모델들의 성능을 비교 분석함으로써 우리가 설계한 모델이 실제로도 꽤 좋은 성능을 보였음을 확인할 수 있었습니다.

추후 더 좋은 성능을 내기 위해서는 Learning rate나 Loss function을 다양하게 주든지 또는 Layer를 조금 더 Deep 하게 쌓음으로써 성능을 높일 수 있을 것입니다. 또한 LeNet이나 ResNet50이 아닌 다른 CNN model로 설계해봄으로써 Accuracy를 조금 더 높일 수 있을 것 입니다.

5. References

[1] 김성진, “피부암의 진단과 치료”, https://www.cnuh.com/health/disease.cs;WEB_JSESSIONID=1BFB761C16F6FB81D7C88BA9E4B0AD9E?act=view&infoId=408&searchKeyword=&searchCondition=&pageIndex=17

[2] 이동윤, “피부암 발병 급증, 피부암 예방 및 자가 전신 관찰법 공개”, 2015.06. ,http://www.samsunghospital.com/home/healthInfo/content/contenView.do?CONT_SRC=HOMEPAGE&CONT_SRC_ID=32756&CONT_CLS_CD=001027&CONT_ID=5366

[3] 고려대학교안암병원, “피부암”, http://anam.kumc.or.kr/dept/disease/deptDiseaseInfoView.do?BNO=435&cPage=&DP_CODE=AADM&MENU_ID=004005

[4] 위키피디아, “피부암”, https://ko.wikipedia.org/wiki/%ED%94%BC%EB%B6%80%EC%95%94

[5] Mathworks, “컨볼루션 뉴럴 네트워크란?”, https://kr.mathworks.com/solutions/deep-learning/convolutional-neural-network.html